GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:44:29 ; Search time 43.01 Seconds
(without alignments)
470.598 Million cell updates/sec

Title: US-09-824-178-3
Perfect score: 605
Sequence: 1 MSQKQEEENPAEETGEEKQD.....PTPQDLPQRKSSLVTSKLAG 117

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	605	100.0	117	4	Q9NRZ0	Q9nrz0 homo sapien
2	605	100.0	121	4	O43768	O43768 homo sapien
3	602	99.5	121	6	O97976	O97976 sus scrofa
4	594	98.2	121	11	Q9Z0G1	Q9z0g1 mus musculu
5	590	97.5	117	11	Q9CQZ9	Q9cqz9 mus musculu
6	400	66.1	110	6	P79322	P79322 sus scrofa
7	226	37.4	119	5	Q9VUB8	Q9vub8 drosophila
8	128.5	21.2	174	5	Q9XU56	Q9xu56 caenorhabdi
9	122	20.2	139	3	P79058	P79058 schizosacch
10	100.5	16.6	417	2	Q9LAY3	Q9lay3 streptococc
11	100	16.5	605	4	Q9NZ53	Q9nz53 homo sapien
12	95.5	15.8	406	2	Q9LAZ0	Q9laz0 streptococc
13	95.5	15.8	842	16	Q98Q49	Q98q49 mycoplasma
14	94.5	15.6	394	2	Q9LAY6	Q9lay6 streptococc
15	94.5	15.6	395	2	Q9LAZ1	Q9laz1 streptococc
16	91.5	15.1	2664	5	Q26033	Q26033 plasmodium
17	91.5	15.1	5120	13	Q9PU36	Q9pu36 gallus gall
18	88	14.5	990	4	Q9NYF9	Q9nyf9 homo sapien
19	88	14.5	1270	4	Q96JH2	Q96jh2 homo sapien
20	87.5	14.5	1751	5	Q9VR59	Q9vr59 drosophila
21	86	14.2	399	10	Q94IK2	Q94ik2 solanum tub
22	85	14.0	775	10	Q9SKL9	Q9skl9 arabidopsis
23	84	13.9	969	5	Q9VLL1	Q9vll1 drosophila
24	84	13.9	969	5	Q95R51	Q95r51 drosophila
25	84	13.9	1165	4	Q9UPN4	Q9upn4 homo sapien
26	84	13.9	2242	4	Q9P2P9	Q9p2p9 homo sapien
27	83.5	13.8	281	5	Q9VYC3	Q9vyc3 drosophila
28	83	13.7	131	3	Q9P305	Q9p305 saccharomyc
29	83	13.7	206	11	Q9CUW6	Q9cuw6 mus musculu
30	83	13.7	487	13	O73793	O73793 serinus can
31	83	13.7	565	5	Q960P3	Q960p3 drosophila
32	83	13.7	619	2	Q54972	Q54972 streptococc
33	83	13.7	735	5	Q9U0D5	Q9u0d5 tetrahymena
34	83	13.7	763	10	Q9SUA1	Q9sua1 arabidopsis
35	83	13.7	792	10	Q9SKM3	Q9skm3 arabidopsis
36	83	13.7	2274	5	Q9VYU0	Q9vyu0 drosophila
37	82.5	13.6	795	10	Q9XGD3	Q9xgd3 arabidopsis
38	82	13.6	2768	5	Q9VC00	Q9vc00 drosophila
39	81.5	13.5	646	10	O64877	O64877 arabidopsis
40	81.5	13.5	789	10	Q9SI46	Q9si46 arabidopsis
41	81.5	13.5	893	4	Q9Y5T0	Q9y5t0 homo sapien
42	81.5	13.5	2817	4	Q96P79	Q96p79 homo sapien
43	81	13.4	474	4	Q96CL1	Q96cl1 homo sapien
44	81	13.4	695	13	Q91550	Q91550 xenopus lae
45	81	13.4	730	2	Q9KK47	Q9kk47 streptococc

ALIGNMENTS

RESULT 1
Q9NRZ0
ID Q9NRZ0 PRELIMINARY; PRT; 117 AA.
AC Q9NRZ0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE BETA ENDOSULFINE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Scott V.E.S., Roch J.-M., Davis-Taber R.A., Molinari E.J.,
RA Whiteaker K.L., Gopalakrishnan M., Idler K., Sullivan J.P.;
RT "Cloning and molecular characterization of two isoforms of human
RT endosulfine.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF157510; AAF80341.1; -.
SQ SEQUENCE 117 AA; 12976 MW; D17E754219767472 CRC64;

Query Match 100.0%; Score 605; DB 4; Length 117;
Best Local Similarity 100.0%; Pred. No. 8.5e-49;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
```

RESULT 2
O43768
ID O43768 PRELIMINARY; PRT; 121 AA.

```
AC   O43768;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA-ENDOSULFINE.
GN   ENSA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND PHOSPHORYLATION.
RC   TISSUE=BRAIN;
RX   MEDLINE=98318667; PubMed=9653196;
RA   Heron L., Virsolvy A., Peyrollier K., Gribble F.M., Le Cam A.,
RA   Ashcroft F.M., Bataille D.;
RT   "Human alpha-endosulfine, a possible regulator of sulfonylurea-
RT   sensitive K(ATP) channel: molecular cloning, expression and biological
RT   properties.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:8387-8391(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99408499; PubMed=10480622;
RA   Heron L., Virsolvy A., Apiou F., Le Cam A., Bataille D.;
RT   "Isolation, characterization, and chromosomal localization of the
RT   human ENSA gene that encodes alpha-endosulfine, a regulator of beta-
RT   cell K(ATP) channels.";
RL   Diabetes 48:1873-1876(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Zhang Q., Fu G., Wu J., Zhou J., Ye M., Shen Y., Kan L., He K., Gu B.,
RA   Chen S., Mao M., Chen Z.;
RT   "Human alpha endosulfine gene.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Scott V.E.S., Roch J.-M., Davis-Taber R.A., Molinari E.J.,
RA   Whiteaker K.L., Gopalakrishnan M., Idler K., Sullivan J.P.;
RT   "Cloning and molecular characterization of two isoforms of human
RT   endosulfine.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG;
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ENDOGENOUS LIGAND FOR SULFONYLUREA RECEPTOR. BY
CC       INHIBITING SULFONYLUREA FROM BINDING TO THE RECEPTOR, IT REDUCES
CC       K(ATP) CHANNEL CURRENTS AND THEREBY STIMULATES INSULIN SECRETION
CC       (PROBABLE).
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGH LEVELS IN MUSCLE
CC       AND BRAIN AND LOWER LEVELS IN THE PANCREAS.
CC   -!- PTM: PHOSPHORYLATED BY PROTEIN KINASE A.
CC   -!- SIMILARITY: STRONG, TO ARPP-19.
DR   EMBL; X99906; CAA68180.1; -.
DR   EMBL; AJ010966; CAB65125.1; -.
DR   EMBL; AF067170; AAD32454.1; -.
DR   EMBL; AF157509; AAF80340.1; -.
DR   EMBL; BC004461; AAH04461.1; -.
DR   EMBL; BC000436; AAH00436.1; -.
DR   MIM; 603061; -.
KW   Phosphorylation.
SQ   SEQUENCE   121 AA;  13389 MW;  7C76315AA17E7542 CRC64;

  Query Match             100.0%;  Score 605;  DB 4;  Length 121;
  Best Local Similarity   100.0%;  Pred. No. 8.9e-49;
  Matches  117;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117


RESULT   3
O97976
ID   O97976      PRELIMINARY;      PRT;   121 AA.
AC   O97976; Q95105;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA-ENDOSULFINE.
OS   Sus scrofa (Pig), and
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823, 9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Pig, and Bovine; TISSUE=BRAIN;
RA   Bataille D.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 25-101 FROM N.A.
RC   SPECIES=Bovine; TISSUE=CAUDATE;
RX   MEDLINE=96268722; PubMed=8687439;
RA   Peyrollier K., Heron L., Virsolvy-Vergine A., Le Cam A., Bataille D.;
RT   "Alpha endosulfine is a novel molecule, structually related to a
RT   family of phosphoproteins.";
RL   Biochem. Biophys. Res. Commun. 223:583-586(1996).
RN   [3]
RP   SEQUENCE OF 25-36; 39-57; 59-74 AND 81-106, AND CHARACTERIZATION.
RC   SPECIES=Pig; TISSUE=BRAIN;
RX   MEDLINE=96218590; PubMed=8635664;
RA   Virsolvy-Vergine A., Salazar G., Sillard R., Denoroy L., Mutt V.,
RA   Bataille D.;
RT   "Endosulfine, endogenous ligand for the sulfonylurea receptor:
RT   isolation and partial structural determination of the alpha form.";
RL   Diabetologia 39:135-141(1996).
CC   -!- FUNCTION: ENDOGENOUS LIGAND FOR SULFONYLUREA RECEPTOR. BY
CC       INHIBITING SULFONYLUREA FROM BINDING TO THE RECEPTOR, IT REDUCES
CC       K(ATP) CHANNEL CURRENTS AND THEREBY STIMULATES INSULIN SECRETION
CC       (PROBABLE).
CC   -!- PTM: PHOSPHORYLATED BY PROTEIN KINASE A (BY SIMILARITY).
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- MASS SPECTROMETRY: MW=13196; METHOD=ELECTROSPRAY.
CC   -!- SIMILARITY: STRONG, TO ARPP-19.
DR   EMBL; AJ005987; CAA06801.1; -.
DR   EMBL; AJ005986; CAA06800.1; -.
DR   EMBL; X95958; CAA65196.1; -.
KW   Phosphorylation.
FT   VARIANT      31     31       K -> R.
SQ   SEQUENCE   121 AA;  13361 MW;  A84629D3B9EFFC5A CRC64;

  Query Match             99.5%;  Score 602;  DB 6;  Length 121;
  Best Local Similarity   99.1%;  Pred. No. 1.7e-48;
  Matches  116;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           |||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db       1 MSQKQEEENPAEETGEEKQDTQEKEGILPEKAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117


RESULT   4
Q9Z0G1
ID   Q9Z0G1      PRELIMINARY;      PRT;   121 AA.
```

```
AC   Q9Z0G1;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA-ENDOSULFINE.
GN   ENSA.
OS   Mus musculus (Mouse), and
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090, 10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Mouse, and Rat; TISSUE=BRAIN;
RA   Bataille D.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Mouse; STRAIN=C57BL/6J; TISSUE=TESTIS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- FUNCTION: ENDOGENOUS LIGAND FOR SULFONYLUREA RECEPTOR. BY
CC       INHIBITING SULFONYLUREA FROM BINDING TO THE RECEPTOR, IT REDUCES
CC       K(ATP) CHANNEL CURRENTS AND THEREBY STIMULATES INSULIN SECRETION
CC       (BY SIMILARITY).
CC   -!- PTM: PHOSPHORYLATED BY PROTEIN KINASE A (BY SIMILARITY).
CC   -!- SIMILARITY: STRONG, TO ARPP-19.
DR   EMBL; AJ005985; CAA06799.1; -.
DR   EMBL; AJ005984; CAA06798.1; -.
DR   EMBL; AK006149; BAB24433.1; -.
DR   MGD; MGI:1891189; Ensa.
KW   Phosphorylation.
SQ   SEQUENCE   121 AA;  13335 MW;  A84632C9085FFC5A CRC64;
```

X bad date

```
  Query Match             98.2%;  Score 594;  DB 11;  Length 121;
  Best Local Similarity   98.3%;  Pred. No. 9.2e-48;
  Matches  115;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db       1 MSQKQEEENPAEETGEEKQDTQEKEGILPEKAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      61 GQKYFDSGDYNMAKAKMKNKQLPSAGADKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
```

```
RESULT   5
Q9CQZ9
ID   Q9CQZ9      PRELIMINARY;      PRT;   117 AA.
AC   Q9CQZ9;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   2610007F17RIK PROTEIN.
GN   ENSA OR 2610007F17RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK019170; BAB31584.1; -.
DR   EMBL; AK011334; BAB27552.1; -.
DR   MGD; MGI:1891189; Ensa.
SQ   SEQUENCE   117 AA;  12922 MW;  3AD5D4F0B07898C8 CRC64;
```

X bad date

```
  Query Match             97.5%;  Score 590;  DB 11;  Length 117;
  Best Local Similarity   97.4%;  Pred. No. 2.1e-47;
  Matches  114;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           ||||||||||||||||||||||:||||||:||||||||||||||||||||||||||||||
Db       1 MSQKQEEENPAEETGEEKQDTQEEEGILPEKAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      61 GQKYFDSGDYNMAKAKMKNKQLPSAGADKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
```

```
RESULT   6
P79322
ID   P79322      PRELIMINARY;      PRT;   110 AA.
AC   P79322;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CAMP-REGULATED PHOSPHOPROTEIN 19K (FRAGMENT).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SMALL INTESTINE;
RA   Winteroe A.K., Fredholm M.;
RT   "Evaluation and characterization of a porcine small intestine cDNA
RT   library.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z84165; CAB06328.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   110 AA;  12189 MW;  DF42271D5CC114D3 CRC64;
```

```
  Query Match             66.1%;  Score 400;  DB 6;  Length 110;
  Best Local Similarity   77.2%;  Pred. No. 7.7e-30;
  Matches   78;  Conservative    6;  Mismatches   17;  Indels    0;  Gaps    0;

Qy      17 EKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAK 76
           |:|   | :   ||:|||||||||:|| |||||||||||  ||||||||||||||||||||
Db       5 EEQKEMEDKVTSPEKAEEAKLKARYPHLGQKPGGSDFLRKRLQKGQKYFDSGDYNMAKAK 64

Qy      77 MKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           | |||||:| |||  |||||||  :|||||| ||| |||||
Db      65 MXNKQLPTAAPDKTXVTGDHIPXHRDLPQRKPSLVASKLAG 105


RESULT   7
Q9VUB8
ID   Q9VUB8      PRELIMINARY;      PRT;   119 AA.
AC   Q9VUB8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG6513 PROTEIN.
GN   CG6513.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003536; AAF49770.1; -.
DR   FlyBase; FBgn0036401; CG6513.
SQ   SEQUENCE   119 AA;  12949 MW;  B5B328A1CBF9E69C CRC64;


  Query Match             37.4%;  Score 226;  DB 5;  Length 119;
  Best Local Similarity   43.7%;  Pred. No. 1.1e-13;
  Matches   52;  Conservative   17;  Mismatches   32;  Indels   18;  Gaps    3;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGG-SDFLMKRLQ 59
           ||  :|  |    | :: : |::      |: || |||:||||  : ||| | || ||||
Db       1 MSSAEENSNSPATTPQDTETTEQANLTDLEKIEEEKLKSKYPSGMRVPGGHSAFLQKRLQ 60

Qy      60 KGQKYFDSGDYNMAKAK-------MKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLV 111
           ||||:|||||| ||| |         ||          : ||: ||||: :| ||:|::
Db      61 KGQKFFDSGDYQMAKQKGGGVKQVFANK----------VTTGEAIPTPETVPARKTSII 109


RESULT   8
Q9XU56
ID   Q9XU56      PRELIMINARY;      PRT;   174 AA.
AC   Q9XU56;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   K10C3.2 PROTEIN.
GN   K10C3.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mortimore B.J.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z83235; CAB05775.1; -.
SQ   SEQUENCE   174 AA;  18108 MW;  64EB4E9176F4B36E CRC64;


  Query Match             21.2%;  Score 128.5;  DB 5;  Length 174;
  Best Local Similarity   29.0%;  Pred. No. 0.00018;
  Matches   38;  Conservative   16;  Mismatches   24;  Indels   53;  Gaps    5;

Qy      29 PERAEEAKLKAKYPSLGQKPG--GSDFLMKRLQKGQKYFDSGDYNMAKAK----MKNKQL 82
           ||: :| :|  |  : |: |    | || |:||: :|:|||||| | |:|    : :|
Db      22 PEKQQEQELMGKLAATGKLPARPASSFLQKKLQQ-RKFFDSGDYAMDKSKAGTGLGSKPH 80

Qy      83 PSAG-------------------------------------------PDKNLVTGDHIPTP 100
           | ||                                           | ||     || |
Db      81 PLAGGPPPAAPPVVAQRSPAPAATTPSPSASPISQQTNRPSSDRNSDDDNL----QIPRP 136

Qy     101 QDLPQRKSSLV 111
             :||||:|::
Db     137 DTVPQRKASII 147


RESULT   9
P79058
ID   P79058      PRELIMINARY;      PRT;   139 AA.
AC   P79058;
DT   01-JAN-1999 (TrEMBLrel. 09, Created)
DT   01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HYPOTHETICAL 15.4 KDA PROTEIN C10F6.16 IN CHROMOSOME I.
GN   SPAC10F6.16.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
```

```
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 20-139 FROM N.A.
RA   Kawamukai M.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO YEAST YNL157W.
DR   EMBL; AL009197; CAA15729.1; -.
DR   EMBL; AB001289; BAA19234.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   139 AA;  15441 MW;  D8D9D2A388ADA532 CRC64;


  Query Match             20.2%;  Score 122;  DB 3;  Length 139;
  Best Local Similarity   34.1%;  Pred. No. 0.00056;
  Matches   30;  Conservative   15;  Mismatches   31;  Indels   12;  Gaps    3;

Qy      20 DTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKN 79
           ::::|  :     || ||   |   |: |   | |:::||:|:|||||||| : ||
Db      20 NSEQKVDVAKLSPEEQKLFRLY---GRLPQRKDLLVQKLQQGRKYFDSGDYALNKA---- 72

Qy      80 KQLPSAGPDKNLV-TGDHIPTPQDLPQR 106
                 |  :   |  ||:|  :| |
Db      73 ----GKASDSGITCIGKEIPSPDTIPHR 96


RESULT  10
Q9LAY3
ID   Q9LAY3      PRELIMINARY;      PRT;   417 AA.
AC   Q9LAY3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PSPA (FRAGMENT).
GN   PSPA.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EF10197;
RX   MEDLINE=20448953; PubMed=10992499;
RA   Hollingshead S.K., Becker R., Briles D.E.;
RT   "Diversity of PspA: mosaic genes and evidence for past recombination
RT   in Streptococcus pneumoniae.";
RL   Infect. Immun. 68:5889-5900(2000).
DR   EMBL; AF071812; AAF27708.1; -.
DR   HSSP; P00192; 1APC.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR001854; Ribosomal_L29.
DR   PRINTS; PR01217; PRICHEXTENSN.
FT   NON_TER     417    417
SQ   SEQUENCE   417 AA;  46960 MW;  876EAD3276506EEC CRC64;


  Query Match             16.6%;  Score 100.5;  DB 2;  Length 417;
  Best Local Similarity   29.4%;  Pred. No. 0.19;
  Matches   37;  Conservative   20;  Mismatches   48;  Indels   21;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           |:: || |::  || :  |||      | ||   | :  || |:    |: :|     :|
Db      50 KRDAEN-AKKALEEAKRAQEKYADYQRRIEEKAAKEQQASLEQQEANKDYQLKL----KK 104

Qy      64 YFD----------SGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDL--PQRKSSLV 111
           | |            :   |: | | ||       : :|    :| ||:|   :||| :|
Db     105 YLDGRNLSNSSVLKKEMEEAEKKDKEKQAEFNKIRREIV----VPNPQELEMARRKSEVV 160

Qy     112 TSKLAG 117
```

```
           :| :|
Db     161 KAKESG 166


RESULT  11
Q9NZ53
ID   Q9NZ53      PRELIMINARY;      PRT;   605 AA.
AC   Q9NZ53;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ENDOGLYCAN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=20187625; PubMed=10722749;
RA   Sassetti C., Van Zante A., Rosen S.D.;
RT   "Identification of Endoglycan, a Member of the CD34/Podocalyxin Family
RT   of Sialomucins.";
RL   J. Biol. Chem. 275:9001-9010(2000).
DR   EMBL; AF219137; AAF44629.1; -.
SQ   SEQUENCE   605 AA;  65075 MW;  2F9B8DC51F7FC22A CRC64;


  Query Match             16.5%;  Score 100;  DB 4;  Length 605;
  Best Local Similarity   27.0%;  Pred. No. 0.32;
  Matches   38;  Conservative   21;  Mismatches   46;  Indels   36;  Gaps    6;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERA--EEAKLKAKYPSL---GQKPGGSDFLMKR 57
           :::|||   |:   |||: :|:| :||     ||||  :  :  ||    | || :
Db     165 EEEEEEEEREKEEVEKQEEEEEEELLPVNGSQEEAKPQVRDFSLTSSSQTPGATK----- 219

Qy      58 LQKGQKYFDSGDYNMAKAKMKNKQ-----LPSAGPDKNLVTGDHIPTPQ---------- 101
                :: ||||   :  ::::       |||  |   :  ||   | |
Db     220 ----SRHEDSGDQASSGVEVESSMGPSLLLPSVTP-TTVTPGDQDSTSQEAEATVLPAAG 274

Qy     102 -----DLPQRKSSLVTSKLAG 117
                 : ||  |   |:  ||
Db     275 LGVEFEAPQEASEEATAGAAG 295


RESULT  12
Q9LAZ0
ID   Q9LAZ0      PRELIMINARY;      PRT;   406 AA.
AC   Q9LAZ0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PSPA (FRAGMENT).
GN   PSPA.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DBL6A;
RX   MEDLINE=20448953; PubMed=10992499;
RA   Hollingshead S.K., Becker R., Briles D.E.;
RT   "Diversity of PspA: mosaic genes and evidence for past recombination
RT   in Streptococcus pneumoniae.";
RL   Infect. Immun. 68:5889-5900(2000).
DR   EMBL; AF071805; AAF27701.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR000533; Tropomyosin.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR00194; TROPOMYOSIN.
FT   NON_TER     406    406
SQ   SEQUENCE   406 AA;  45406 MW;  5B1E20CA4D06F052 CRC64;
```

```
  Query Match             15.8%;  Score 95.5;  DB 2;  Length 406;
  Best Local Similarity   28.6%;  Pred. No. 0.54;
  Matches   36;  Conservative   19;  Mismatches   50;  Indels   21;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           |:: || |::  || :  |||      | ||   |  : || |:    |: :|     :|
Db      50 KRDAEN-AKKALEEAKRAQEKYADYQRRIEEKAAKETHASLEQQEANKDYQLKL----KK 104

Qy      64 YFD----------SGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDL--PQRKSSLV 111
           | |            :   |: | | ||       : :|    :| ||:|   :||| :
Db     105 YLDGRNLSNSSVLKKEMEEAEKKDKEKQAEFNKIRREIV----VPNPQELEMARRKSEVA 160

Qy     112 TSKLAG 117
             | :|
Db     161 KPKESG 166


RESULT  13
Q98Q49
ID   Q98Q49      PRELIMINARY;      PRT;   842 AA.
AC   Q98Q49;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   HEAT SHOCK ATP-DEPENDENT PROTEASE (EC 3.4.21.53).
GN   MYPU_5190.
OS   Mycoplasma pulmonis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2107;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UAB CTIP;
RX   MEDLINE=21267165; PubMed=11353084;
RA   Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA   Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA   Blanchard A.;
RT   "The complete genome sequence of the murine respiratory pathogen
RT   Mycoplasma pulmonis.";
RL   Nucleic Acids Res. 29:2145-2153(2001).
DR   EMBL; AL445564; CAC13692.1; -.
DR   MypuList; MYPU_5190; -.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003959; AAA_subfam.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   InterPro; IPR003111; LON.
DR   InterPro; IPR001984; Lon_endopep.
DR   Pfam; PF00004; AAA; 1.
DR   PRINTS; PR00830; ENDOLAPTASE.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00464; LON; 1.
KW   Hydrolase; Complete proteome.
SQ   SEQUENCE   842 AA;  95888 MW;  99A276B296EDBFC2 CRC64;


  Query Match             15.8%;  Score 95.5;  DB 16;  Length 842;
  Best Local Similarity   31.5%;  Pred. No. 1.2;
  Matches   35;  Conservative   19;  Mismatches   48;  Indels    9;  Gaps    4;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQ-KPGGSDFLMKRLQKG 61
           : |: ||  :|  ::    |:|| || |     |:||    ||:       ::  |  |
Db     234 EDQDLENAIDEKMKKNLSKQQKEFILRE-----KMKAIRSELGESNDDEEEYKTKIKDKN 288

Qy      62 QK--YFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSL 110
            |  |  |    :|  : | | : |: || | :| ::| |   || ||:|:
Db     289 NKKIYPKSVRKLIANEESKLKNMMSSSPDAN-ITRNYIDTLMKLPWRKTSI 338


RESULT  14
Q9LAY6
ID   Q9LAY6      PRELIMINARY;      PRT;   394 AA.
AC   Q9LAY6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PSPA (FRAGMENT).
GN   PSPA.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=L81905;
RX   MEDLINE=20448953; PubMed=10992499;
RA   Hollingshead S.K., Becker R., Briles D.E.;
RT   "Diversity of PspA: mosaic genes and evidence for past recombination
RT   in Streptococcus pneumoniae.";
RL   Infect. Immun. 68:5889-5900(2000).
DR   EMBL; AF071809; AAF27705.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR000533; Tropomyosin.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR00194; TROPOMYOSIN.
FT   NON_TER     394    394
SQ   SEQUENCE   394 AA;  42935 MW;  A192F47E12C88FAA CRC64;


  Query Match             15.6%;  Score 94.5;  DB 2;  Length 394;
  Best Local Similarity   28.6%;  Pred. No. 0.64;
  Matches   36;  Conservative   19;  Mismatches   50;  Indels   21;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           |:: || |::  || :  |||      | ||   |    || |:    |: :|     :|
Db      50 KRDAEN-AKKALEEAKRAQEKYADYQRRIEEKAAKETQASLEQQEANKDYQLKL----KK 104

Qy      64 YFD----------SGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDL--PQRKSSLV 111
           | |            :   |: | | ||       : :|    :| ||:|   :||| :
Db     105 YLDGRNLSNSSVLKKEMEEAEKKDKEKQAEFNKIRREIV----VPNPQELEMARRKSEVA 160

Qy     112 TSKLAG 117
             :| :|
Db     161 KAKESG 166


RESULT  15
Q9LAZ1
ID   Q9LAZ1      PRELIMINARY;      PRT;   395 AA.
AC   Q9LAZ1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PSPA (FRAGMENT).
GN   PSPA.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BG9739;
RX   MEDLINE=20448953; PubMed=10992499;
RA   Hollingshead S.K., Becker R., Briles D.E.;
RT   "Diversity of PspA: mosaic genes and evidence for past recombination
RT   in Streptococcus pneumoniae.";
RL   Infect. Immun. 68:5889-5900(2000).
DR   EMBL; AF071804; AAF27700.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR000533; Tropomyosin.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR00194; TROPOMYOSIN.
FT   NON_TER     395    395
SQ   SEQUENCE   395 AA;  42968 MW;  67FDECAC41DB7F95 CRC64;
```

```
  Query Match             15.6%;  Score 94.5;  DB 2;  Length 395;
  Best Local Similarity   28.6%;  Pred. No. 0.64;
  Matches   36;  Conservative   19;  Mismatches   50;  Indels   21;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           |:: || |::  || :  |||      | ||   |    || |:    |: :|     :|
Db      50 KRDAEN-AKKALEEAKRAQEKYADYQRRIEEKAAKETQASLEQQEANKDYQLKL----KK 104

Qy      64 YFD----------SGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDL--PQRKSSLV 111
           | |          :   |: | | ||      : :|    :| ||:|   :||| :
Db     105 YLDGRNLSNSSVLKKEMEEAEKKDKEKQAEFNKIRREIV----VPNPQELEMARRKSEVA 160

Qy     112 TSKLAG 117
            :| :|
Db     161 KAKESG 166


RESULT  16
Q26033
ID   Q26033      PRELIMINARY;      PRT;  2664 AA.
AC   Q26033;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   VARIANT-SPECIFIC SURFACE PROTEIN.
GN   VAR-2.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FCR3;
RX   MEDLINE=95330813; PubMed=7606788;
RA   Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
RA   Peterson D.S., Ravetch J.A., Wellems T.E.;
RT   "The large diverse gene family var encodes proteins involved in
RT   cytoadherence and antigenic variation of Plasmodium falciparum-
RT   infected erythrocytes.";
RL   Cell 82:89-100(1995).
DR   EMBL; L40609; AAA75398.1; -.
DR   InterPro; IPR004258; PFEMP.
DR   Pfam; PF03011; PFEMP; 2.
SQ   SEQUENCE   2664 AA;  302410 MW;  6EA2468511703091 CRC64;


  Query Match             15.1%;  Score 91.5;  DB 5;  Length 2664;
  Best Local Similarity   26.8%;  Pred. No. 11;
  Matches   34;  Conservative   16;  Mismatches   30;  Indels   47;  Gaps    5;

Qy       6 EEENPAEETG-----EEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           ||||| |: |     ::: :  :| | | |: :| | : : |:  :: ||:
Db    2086 EEENPVEQPGFCPTPQQEPEPDDKCGKLEEKKDEKKEQPEQPA--EEDGGA--------- 2134

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGP------DKNLVTGDHIPTPQDLPQRK-----SS 109
                              :||  |      ||  |    || ||: |          |
Db    2135 -------------------IVPSGPPGSEPEADKGPVKPAEIPKPQEPPDLSHPAVIPS 2174

Qy     110 LVTSKLA 116
           |||| ||
Db    2175 LVTSTLA 2181


RESULT  17
Q9PU36
ID   Q9PU36      PRELIMINARY;      PRT;  5120 AA.
AC   Q9PU36;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ACZONIN (FRAGMENT).
GN   ACZ.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99439764; PubMed=10508862;
RA   Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA   Kilimann M.W.;
RT   "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT   zones, shares homology regions with rim and bassoon and binds
RT   profilin.";
RL   J. Cell Biol. 147:151-162(1999).
DR   EMBL; Y19187; CAB60725.1; -.
DR   HSSP; P04410; 1A25.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR001565; Synaptotagmin.
DR   Pfam; PF00168; C2; 2.
DR   Pfam; PF00595; PDZ; 1.
DR   PRINTS; PR00360; C2DOMAIN.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR00399; SYNAPTOTAGMN.
DR   SMART; SM00239; C2; 2.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS00499; C2_DOMAIN_1; UNKNOWN_1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 2.
DR   PROSITE; PS50106; PDZ; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   5120 AA;  560751 MW;  A658D9891B65B412 CRC64;


  Query Match             15.1%;  Score 91.5;  DB 13;  Length 5120;
  Best Local Similarity   21.8%;  Pred. No. 23;
  Matches   32;  Conservative   26;  Mismatches   48;  Indels   41;  Gaps    3;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGIL---------PERAEEAKLKAKYPSLGQKPGGSDF 53
           || | | | ::  |:||:  :|| :          ||:||| |       | :    ||
Db    1194 QKVEVEAPCDKLHEKKQEDVKKEDLTTGIPQMVSKPEKAEEEKTPVPVSRLPR----SDH 1249

Qy      54 LMKRLQKGQKYFDSGDYNMAKAKMK---------------------------NKQLPSA 85
           :    :| :|  |  | : :: :                             | ||
Db    1250 VEAVREKIEKEDDKSDTSSSQQQKSPQGLSDTGYSSDGISSSLGEIPSHIPSDEKDLPRE 1309

Qy      86 GPDKNLVTGDHIPTPQDLPQRKSSLVT 112
             |: :: :  |:| || : :|::::
Db    1310 PSQKDTISQESPPSPSDLAKLESTVLS 1336


RESULT  18
Q9NYF9
ID   Q9NYF9      PRELIMINARY;      PRT;  990 AA.
AC   Q9NYF9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   GOLGI MATRIX PROTEIN GM130.
GN   GOLGA2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   Weide T.B., Bayer M., Barnekow A.;
RT   "Human golgi matrix protein 130, huGM130, a protein highly homologous
RT   to GM130 from rat.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF248953; AAF65550.1; -.
KW   Matrix protein.
```

```
SQ   SEQUENCE   990 AA;  111658 MW;  2380FBFD11B8DAF5 CRC64;


  Query Match             14.5%;  Score 88;  DB 4;  Length 990;
  Best Local Similarity   25.9%;  Pred. No. 7.4;
  Matches   37;  Conservative   22;  Mismatches   50;  Indels   34;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPER------------AEEAKLKAKYP----SLGQK 47
           ||: :   ::  |||::  :|:| | |:            :|:|:|:        :  ||
Db     176 KQQNQEITDQLEEEKKECHQKQGALREQLQVHIQTIGILVSEKAELQTALAHTQHAARQK 235

Qy      48 PGGSDFLMKRLQKGQKYFDSGDYNMA---------KAKMKNKQLPSAGPDKNLVTGDHIP 98
           | |: |  |||  ::    |:   |          ||   ||:|        |    :
Db     236 EGESEDLASRLQYSRRRV--GELERALSAVSTQQKKADRYNKELTKERDALRLELYKNTQ 293

Qy      99 TPQDLPQRKSS-------LVTSK 114
           : :|| | ||         ||| |
Db     294 SNEDLKQEKSELEEKLRVLVTEK 316


RESULT  19
Q96JH2
ID   Q96JH2      PRELIMINARY;      PRT;  1270 AA.
AC   Q96JH2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA1855 PROTEIN (FRAGMENT).
GN   KIAA1855.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=21245130; PubMed=11347906;
RA   Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XX.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large Proteins in vitro.";
RL   DNA Res. 8:85-95(2001).
DR   EMBL; AB058758; BAB47484.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   1270 AA;  137310 MW;  61D4AB45414E3FEC CRC64;


  Query Match             14.5%;  Score 88;  DB 4;  Length 1270;
  Best Local Similarity   30.0%;  Pred. No. 9.8;
  Matches   36;  Conservative   18;  Mismatches   52;  Indels   14;  Gaps    4;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPG----GSDFLMKRLQ 59
           ::|||   ||  ||::: :|:|    |  | |   |    || : |    ||:    | |
Db     691 EEEEEEDEEEEEEEEEEEEEEEEEEEEEEEAAAAVALGEVLGPRSGSSSEGSERSTDRSQ 750

Qy      60 KGQK---YFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLA 116
           :|      |    :  :| | :  |      | ||    : :| |:   | | ||::||
Db     751 EGAPSTLLADDQKESRGRASMADGDLEPEEGSKTLV----LVSPGDM---KKSPVTAELA 803


RESULT  20
Q9VR59
ID   Q9VR59      PRELIMINARY;      PRT;  1751 AA.
AC   Q9VR59;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG14616 PROTEIN.
GN   CG14616.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003574; AAF50947.1; -.
DR   FlyBase; FBgn0031192; CG14616.
DR   InterPro; IPR000560; His_acid_phosphtse.
DR   PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; UNKNOWN_1.
SQ   SEQUENCE   1751 AA;  194376 MW;  4556EA0BE7B2C8E4 CRC64;


  Query Match             14.5%;  Score 87.5;  DB 5;  Length 1751;
  Best Local Similarity   26.3%;  Pred. No. 16;
  Matches   35;  Conservative   18;  Mismatches   57;  Indels   23;  Gaps    4;

Qy       3 QKQEEENPA-------EETGEEKQDTQEKEGILPERAEEAKLKAKYPSLG---QKPGGSD 52
           |  :| |||          :|||::   :| |   :|: :   :    |: |    :
Db     976 QSSDESNPARIEEENDSNSGEEREG--KKRGTSGQRSTDRSAERISPAFGFNRLELRSKQ 1033

Qy      53 FLMKRLQKGQKYFDSGDYNMAKAK----------MKNKQLPSAGPDKNLVTGDHIPTPQ 101
           |  | :  |  :  ||   |  ||           :|:||    ||   || |  |  :
Db    1034 FKSKPIPIGAHHTVSGHEAMDLAKRLNEELASHQQQQNQQLRPISPDIRAVTPDCEPRSR 1093

Qy     102 DLPQRKSSLVTSK 114
             || :| | :|
Db    1094 SFEQRPTSGVCAK 1106


RESULT  21
Q94IK2
ID   Q94IK2      PRELIMINARY;      PRT;   399 AA.
AC   Q94IK2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
```

```
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   STOREKEEPER PROTEIN.
GN   STK.
OS   Solanum tuberosum (Potato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4113;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. DESIREE;
RA   Zourelidou M., de Torres Zabala M., Smith C., Bevan M.W.;
RT   "STORE KEEPER (STK) regulates the expression of the potato storage
RT   protein patatin and is the first member of a plant-specific family of
RT   DNA-binding proteins.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. DESIREE;
RA   Zourelidou M.;
RL   Thesis (1997), University of East Anglia, Norwich, Norfolk,
RL   United Kingdom.
DR   EMBL; AJ314612; CAC39398.1; -.
SQ   SEQUENCE   399 AA;  44232 MW;  A118D95465799F1E CRC64;


  Query Match             14.2%;  Score 86;  DB 10;  Length 399;
  Best Local Similarity   27.3%;  Pred. No. 4;
  Matches   27;  Conservative   13;  Mismatches   37;  Indels   22;  Gaps    2;

Qy       2 SQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKG 61
           ||::||:   || |||:   : :|   |: |     |     | | |          ||
Db      28 SQEEEEQQSREEEGEEESGEETEEDEEPKTAHPVVKKPISQKLVQTP----------QKP 77

Qy      62 QKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTP 100
           |   :||  |           :| |    :|: :|:|.
Db      78 QFSSESGSEN-----------GSGSDSEAESGNSLPSP 104


RESULT  22
Q9SKL9
ID   Q9SKL9      PRELIMINARY;      PRT;   775 AA.
AC   Q9SKL9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AT2G15190 PROTEIN.
GN   AT2G15190.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006298; AAD25573.1; -.
DR   InterPro; IPR003653; SUMO_protease.
DR   Pfam; PF02902; Ulp1_C; 1.
SQ   SEQUENCE   775 AA;  88957 MW;  CB1F20F50454A31F CRC64;


  Query Match             14.0%;  Score 85;  DB 10;  Length 775;
  Best Local Similarity   20.6%;  Pred. No. 11;
  Matches   32;  Conservative   24;  Mismatches   43;  Indels   56;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAE------EAKLKAKYPSLGQKPGGSDFLMK- 56
           :::||   |: | |:|   |:|| |  :||      |  ::      |  ||  :   :
Db     375 EEDEERKQEDEGVERQPEAEEEGGLERKAENDNESFEDSIREPNTQYGSYPGDDENTQRD 434

Qy      57 --------------------------------RLQKGQKYFDSGDYNMAKAKMKNKQL-P 83
                                            :|| :|       ::: : :||:: |
Db     435 VGDEVAEESSKDKSPTPRSSTPNFNILSEESLDVQKDKK-------RVSRGRNENKRVKP 487

Qy      84 SAGPDKNLVTGDHIPTPQ---------DLPQRKSS 109
           |   : || |   :|  :         |:| || |
Db     488 SVHAEDNLKTRKQVPRKRQKQVDSADVDVPTRKES 522


RESULT  23
Q9VLL1
ID   Q9VLL1      PRELIMINARY;      PRT;   969 AA.
AC   Q9VLL1; Q9XZU3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   D12 PROTEIN.
GN   D12 OR CG13400.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
```

```
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE OF 233-969 FROM N.A.
RA   Hovemann B.T., Heiermann R., Richardt A.;
RT   "Transcriptional activity at the chromosomal locus 29C.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003621; AAF52676.1; -.
DR   EMBL; AJ242723; CAB44307.1; -.
DR   FlyBase; FBgn0027490; D12.
KW   Hypothetical protein.
SQ   SEQUENCE   969 AA;  111119 MW;  6D4EAA9A51367867 CRC64;


  Query Match             13.9%;  Score 84;  DB 5;  Length 969;
  Best Local Similarity   27.4%;  Pred. No. 17;
  Matches   32;  Conservative   20;  Mismatches   51;  Indels   14;  Gaps    5;

Qy       1 MSQKQEEEN--PAEETGEEK-QDTQEKEGILP-----ERAEEAKLKAKYPSLGQKPGGSD 52
           :| : |||    ||  |: |   |:: :|:||     :|  |          : ||
Db      27 VSSQDEEETCAEAESPGQRKWTKTEQSDGMLPLSDKFQRIRELLRLEFQREISQKVEQLA 86

Qy      53 FLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPD-----KNLVTGDHIPTPQDLP 104
           : :|| :|::  |   | :     : :|:|  | |       : : || |  || ||
Db      87 EIDRRLLQGRQLLDRLRYQVVSEYYRKQQVPLTGADIAKVRGDSLFGDDIAAPQ-LP 142


RESULT  24
Q95R51
ID   Q95R51      PRELIMINARY;      PRT;   969 AA.
AC   Q95R51;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SD07884P.
GN   D12.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY061612; AAL29160.1; -.
SQ   SEQUENCE   969 AA;  111048 MW;  567E1A22FA17A82F CRC64;


  Query Match             13.9%;  Score 84;  DB 5;  Length 969;
  Best Local Similarity   27.4%;  Pred. No. 17;
  Matches   32;  Conservative   20;  Mismatches   51;  Indels   14;  Gaps    5;

Qy       1 MSQKQEEEN--PAEETGEEK-QDTQEKEGILP-----ERAEEAKLKAKYPSLGQKPGGSD 52
           :| : |||    ||  |: |   |:: :|:||     :|  |          : ||
Db      27 VSSQDEEETCAEAESPGQRKWTKTEQSDGMLPLSDKFQRIRELLRLEFQREISQKVEQLA 86

Qy      53 FLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPD-----KNLVTGDHIPTPQDLP 104
           : :|| :|::  |   | :     : :|:|  | |       : : || |  || ||
Db      87 EIDRRLLQGRQLLDRLRYQVVSEYYRKQQVPLTGADIAKVRGDSLFGDDIAAPQ-LP 142


RESULT  25
Q9UPN4
ID   Q9UPN4      PRELIMINARY;      PRT;  1165 AA.
AC   Q9UPN4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA1118 PROTEIN (FRAGMENT).
GN   KIAA1118.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99397452; PubMed=10470851;
RA   Kikuno R., Nagase T., Ishikawa K., Hirosawa M., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIV.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:197-205(1999).
DR   EMBL; AB029041; BAA83070.1; -.
DR   InterPro; IPR000048; IQ.
DR   InterPro; IPR001978; Troponin.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF00992; Troponin; 1.
DR   SMART; SM00015; IQ; 1.
FT   NON_TER       1       1
SQ   SEQUENCE   1165 AA;  130214 MW;  D5517BFB825FDF1A CRC64;


  Query Match             13.9%;  Score 84;  DB 4;  Length 1165;
  Best Local Similarity   23.2%;  Pred. No. 21;
  Matches   32;  Conservative   18;  Mismatches   46;  Indels   42;  Gaps    5;

Qy       5 QEEENPAEETGEEK---------QDTQEKEGILPERAEEAK--------------LKAKY 41
           |::|    :  |||          |: |:|  :  ::|  |:              ::
Db     400 QQKEAARRKAREEKARQARRAAIQELQQKRALRAQKASTAERGPPENPRETRVPGMRQPA 459

Qy      42 PSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGP-DKNLVTGDHIPTP 100
             |   |||:                  :   ||      ||:||| |: | | |  | |
Db     460 QELSPTPGGT----------------AHQALKANNAGGGLPAAGPGDRCLPTSDSSPEP 502

Qy     101 QDLPQ-RKSSLVTSKLAG 117
           |  |: |   ::    ||
Db     503 QQPPEDRTQDVLAQDAAG 520


RESULT  26
Q9P2P9
ID   Q9P2P9      PRELIMINARY;      PRT;  2242 AA.
AC   Q9P2P9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA1297 PROTEIN (FRAGMENT).
GN   KIAA1297.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=20181126; PubMed=10718198;
RA   Nagase T., Kikuno R., Ishikawa K., Hirosawa M., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes.XVI.
RT   The complete sequences of 150 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 7:65-73(2000).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AB037718; BAA92535.1; -.
DR   HSSP; P56276; 1TLK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003961; FN_III.
```

```
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 2.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00410; IG_like; 3.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 2.
KW   ATP-binding; Serine/threonine-protein kinase; Transferase.
FT   NON_TER       1      1
FT   NON_TER    2242   2242
SQ   SEQUENCE   2242 AA;  243491 MW;  497946B526A3561A CRC64;


  Query Match             13.9%;  Score 84;  DB 4;  Length 2242;
  Best Local Similarity   27.4%;  Pred. No. 44;
  Matches   29;  Conservative   16;  Mismatches   43;  Indels   18;  Gaps    4;

Qy       5 QEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKP--GGSDFLMKRLQKGQ 62
           ::|    |||      :::|  | : :||     || ||:|  | |       ::|:
Db     975 EDEALGTPETGAATPMDWQEQGRAPSQDQEAPSPEALPSPGQEPAAGASP------RRGE 1028

Qy      63 KYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKS 108
                 : :       || ||| : |  | |     :||||:|
Db   1029 ---------LRRGSSAESALPRAGP-RELGRGLHKAASVELPQRRS 1064


RESULT  27
Q9VYC3
ID   Q9VYC3      PRELIMINARY;      PRT;   281 AA.
AC   Q9VYC3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG15754 PROTEIN.
GN   CG15754.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003492; AAF48280.1; -.
DR   FlyBase; FBgn0030492; CG15754.
SQ   SEQUENCE   281 AA;  31630 MW;  34F3DD5BD11B5398 CRC64;


  Query Match             13.8%;  Score 83.5;  DB 5;  Length 281;
  Best Local Similarity   36.5%;  Pred. No. 4.6;
  Matches   27;  Conservative   12;  Mismatches   20;  Indels   15;  Gaps    4;

Qy       3 QKQEEENP-----AEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKR 57
             | ||||:|     |||  ||::| :|:|   |:       :   |:| | |||   :::
Db     120 QGQEEESPDQDGEAEEAEEEEDEEEQEEATPQP------QPLPPALPQ-PGG---VLEE 169

Qy      58 LQKGQKYFDSGDYN 71
           |  ||   :  |||
Db     170 LAGGQVDEELEDYN 183


RESULT  28
Q9P305
ID   Q9P305      PRELIMINARY;      PRT;   131 AA.
AC   Q9P305;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   YHR132W-AP.
GN   YHR132W-A.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94378003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Keppler D., Langston Y.,
RA   Latreille P., Louis E., Macri C., Mardis E., Mouser L., Nhan M.,
RA   Rifken L., Riles L., St Peter H., Thornton L., Trevaskis E.,
RA   Vaudin M., Vaughan K., Vignati D., Wilcox L., Willis A., Wilson R.,
RA   Wohldman P., Waterston R.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Fulton L.;
RT   "The sequence of S. cerevisiae cosmid 9315.";
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Jia Y., Cherry J.M.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Cherry J.M.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U10398; AAF28366.1; -.
DR   SGD; S0007496; YHR132W-A.
```

```
SQ   SEQUENCE   131 AA;  14395 MW;  A6B6D9298D9FD013 CRC64;


  Query Match             13.7%;  Score 83;  DB 3;  Length 131;
  Best Local Similarity   32.4%;  Pred. No. 2.1;
  Matches   24;  Conservative    9;  Mismatches   29;  Indels   12;  Gaps    3;

Qy      24 KEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKM------ 77
           |||:   :    :||  |   |: |   | |  ::|  | |||||| : || :
Db      22 KEGVDLSKLSPQELKL-YKMYGKLPSKKDLLRHKMQDRQ-YFDSGDYALKKAGVIKSDDV 79

Qy      78 ----KNKQLPSAGP 87
                :  ||   |
Db      80 IVNNSSNNLPVTNP 93


RESULT  29
Q9CUW6
ID   Q9CUW6      PRELIMINARY;      PRT;   206 AA.
AC   Q9CUW6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2900056O04RIK PROTEIN (FRAGMENT).
GN   2900056O04RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=HIPPOCAMPUS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK013708; BAB28962.1; -.
DR   MGD; MGI:1920234; 2900056O04Rik.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   InterPro; IPR000690; ZnF_matrin.
DR   InterPro; IPR003604; ZnF_U1.
DR   SMART; SM00451; ZnF_U1; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
FT   NON_TER      1      1
SQ   SEQUENCE   206 AA;  23233 MW;  4316F75C83BC8F72 CRC64;


  Query Match             13.7%;  Score 83;  DB 11;  Length 206;
  Best Local Similarity   24.8%;  Pred. No. 3.6;
  Matches   35;  Conservative   16;  Mismatches   58;  Indels   32;  Gaps    4;

Qy       2 SQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPG---GSDFLMKRL 58
           | ::|:|:  ||  || :   |:|     :  |   :    |    |   | |||:  :
Db      47 SGQEEDEDDDEEEEEEGEIEAEEEFCKQVKPRETSSEQGKGSETYNPNTAYGEDFLVPVM 106

Qy      59 ----QKGQKYFDSGD------------YNMAKAKMKNKQLPSAGPDKN---------- 90
```

```
              |   |::||                |: | | ||   |   |
Db     107 GYVCQICHKFYDSNSELRFSHCKSLAHFENLQKYKAKNPTPPPTRPVSRKCAINARNALT 166

Qy      91 -LVTGDHIPTPQDLPQRKSSL 110
            | |  | |:||| :  | :
Db     167 ALFTSSHQPSPQDTVKMPSKV 187


RESULT  30
O73793
ID   O73793      PRELIMINARY;      PRT;   487 AA.
AC   O73793;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEUROFILAMENT MEDIUM SUBUNIT.
OS   Serinus canaria (Canary).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Passeroidea;
OC   Fringillidae; Carduelinae; Serinus.
OX   NCBI_TaxID=9135;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Siepka S.M., Clayton D.F.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR   EMBL; AF053711; AAC06245.1; -.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Coiled coil; Intermediate filament.
SQ   SEQUENCE   487 AA;  53397 MW;  470F20B0892DDF43 CRC64;


  Query Match             13.7%;  Score 83;  DB 13;  Length 487;
  Best Local Similarity   25.9%;  Pred. No. 9.5;
  Matches   37;  Conservative   17;  Mismatches   59;  Indels   30;  Gaps    5;

Qy       3 QKQEEE---------NPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQK------ 47
           :|:|||          : ||| | ||:: :|||       |||: | |    | |
Db     143 EKEEEEAEEEEAAKSDAAEEGGSEKEEIEEKEEGEEAEEEEAEAKGKAEEAGAKVEKVKS 202

Qy      48 ---------PGGSDFL--MKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNL---VT 93
                   |  |      | :||              |:   |:  :| |:|     ||
Db     203 PPAKSPPKSPPKSPVTEQAKAVQKAAAEVGKDQKAEKAAEKAAKEEKAASPEKPATPKVT 262

Qy      94 GDHIP-TPQDLPQRKSSLVTSKL 115
           | ||:  |  : ::   |:
Db     263 SPEKPATPEKPPTPEKAITPEKV 285


RESULT  31
Q960P3
ID   Q960P3      PRELIMINARY;      PRT;   565 AA.
AC   Q960P3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LD41932P.
GN   CG1905.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Y, CN BW SP;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
```

```
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY051941; AAK93365.1; -.
SQ   SEQUENCE   565 AA;  63407 MW;  3F53DE0AECB8701B CRC64;


  Query Match             13.7%;  Score 83;  DB 5;  Length 565;
  Best Local Similarity   30.6%;  Pred. No. 11;
  Matches   38;  Conservative   15;  Mismatches   37;  Indels   34;  Gaps    7;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGIL--PERAE-------EAKLKAKYPSLGQKPGGS 51
           : :|: ||    || ||:|  :|:|  |   ||||       | ||| |   | :|
Db      66 LKEKEREEKLKRET-EERQREKEREEKLKEKERAEKLKDLEKEVKLKEKEEQLKEK---- 120

Qy      52 DFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPS---------AGPDKNLV--TGDHIPTP 100
           :  :|  :| :|          | | | | | |         : | : :|  |   :|
Db     121 EKELKLKEKKEK---------DKVKEKEKSLESEKLLISATVSNPWRRVVEDTPPKLPAV 171

Qy     101 QDLP 104
           || |
Db     172 QDYP 175


RESULT  32
Q54972
ID   Q54972      PRELIMINARY;      PRT;   619 AA.
AC   Q54972;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PNEUMOCOCCAL SURFACE PROTEIN A PRECURSOR.
GN   PSPA.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yother J., Briles D.E.;
RT   "Structural and evolutionary properties of PspA, a surface protein of
RT   Streptococcus pneumoniae, as revealed by sequence analysis.";
RL   Submitted (MAR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92105030; PubMed=1729249;
RA   Yother J., Briles D.E.;
RT   "Structural properties and evolutionary relationships of PspA, a
RT   surface protein of Streptococcus pneumoniae, as revealed by sequence
RT   analysis.";
RL   J. Bacteriol. 174:601-609(1992).
DR   EMBL; M74122; AAA27018.1; -.
DR   HSSP; P00192; 1APC.
DR   InterPro; IPR002479; CW_binding.
DR   InterPro; IPR000566; Lipocln_cytFABP.
DR   Pfam; PF01473; CW_binding_1; 10.
DR   PROSITE; PS00213; LIPOCALIN; UNKNOWN_3.
KW   Signal.
FT   SIGNAL        1     31       POTENTIAL.
FT   CHAIN        32    619       PNEUMOCOCCAL SURFACE PROTEIN A.
SQ   SEQUENCE   619 AA;  68605 MW;  5AA8BDB40C2841CA CRC64;


  Query Match             13.7%;  Score 83;  DB 2;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 13;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
           : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  33
Q9U0D5
ID   Q9U0D5      PRELIMINARY;      PRT;   735 AA.
AC   Q9U0D5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KINESIN-II HOMOLOGUE.
GN   KIN1.
OS   Tetrahymena thermophila.
OC   Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC   Tetrahymenina; Tetrahymena.
OX   NCBI_TaxID=5911;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Brown J.M., Marsala C., Kosoy R., Gaertig J.;
RT   "Kinesin-II is preferentially targeted to assembling cilia and is
RT   required for ciliogenesis and normal cytokinesis in Tetrahymena.";
RL   Mol. Biol. Cell 3081:3081-3096(1999).
DR   EMBL; AJ244020; CAB51556.1; -.
DR   HSSP; P33176; 1BG2.
DR   InterPro; IPR001752; kinesin.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   ATP-binding; Coiled coil; Microtubules; Motor protein.
SQ   SEQUENCE   735 AA;  85037 MW;  8395EDB1BB97C78D CRC64;


  Query Match             13.7%;  Score 83;  DB 5;  Length 735;
  Best Local Similarity   26.5%;  Pred. No. 15;
  Matches   31;  Conservative   21;  Mismatches   35;  Indels   30;  Gaps    4;

Qy       1 MSQKQEEE---------NPAEETGEEKQ-DTQEKEGILPERAEEAKLKAKYPSLGQKPGG 50
           : || |:|          |  ::  |:|:   :||: :| :  :: : : |     :|
Db     405 LQQKLEQEKSEVEKKYQNEVKKIEEDKKLQEKEKQQLLQQLMDKEERERQKQREQQEK--- 461

Qy      51 SDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRK 107
             |:|:||| |:    ||  | ||  | ::|  |                |:| |||
Db     462 ---LLKKLQKMQEKVIQGDEIMQKALEKERELNKA-------------QQELQQRK 501


RESULT  34
Q9SUA1
ID   Q9SUA1      PRELIMINARY;      PRT;   763 AA.
AC   Q9SUA1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYPOTHETICAL 85.2 KDA PROTEIN.
GN   T15N24.80 OR AT4G26630.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Zimmermann W., Grueneisen A., Wambutt R., Bancroft I.,
RA   Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Zimmermann W., Grueneisen A., Wambutt R., Kalicki J., Wohldmann P.,
RA   Smith A., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
```

```
RN   [4]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL078465; CAB43859.1; -.
DR   EMBL; AL161565; CAB79518.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   763 AA;  85249 MW;  D67F4F5BD4DB480B CRC64;


  Query Match             13.7%;  Score 83;  DB 10;  Length 763;
  Best Local Similarity   25.0%;  Pred. No. 16;
  Matches   31;  Conservative   25;  Mismatches   36;  Indels   32;  Gaps    6;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGI--LPERAEEAKLKAKYPSLGQKPGGSDFLMK---- 56
           :|:|||   ||  |||:: :|:|    :|:::|:   :|  ||  ::   |:   :
Db     526 EKEEEEKQEEEKAEEKEEKKEEENENGIPDKSED---EAPQPSESEEKDESEEHSEEETT 582

Qy      57 ------RLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSL 110
                 ||  |:|         :  : :||        | :|     | |: : |::||
Db     583 KKKRGSRLSAGKK--------ESAGRARNK--------KAVVAAKSSP-PEKITQKRSSA 625

Qy     111 VTSK 114
               |
Db     626 KRKK 629


RESULT  35
Q9SKM3
ID   Q9SKM3      PRELIMINARY;      PRT;   792 AA.
AC   Q9SKM3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AT2G15140 PROTEIN.
GN   AT2G15140.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006298; AAD25576.1; -.
DR   InterPro; IPR003653; SUMO_protease.
DR   Pfam; PF02902; Ulp1_C; 1.
SQ   SEQUENCE   792 AA;  91035 MW;  D39C5C03CCC28D75 CRC64;


  Query Match             13.7%;  Score 83;  DB 10;  Length 792;
  Best Local Similarity   21.3%;  Pred. No. 17;
  Matches   33;  Conservative   25;  Mismatches   41;  Indels   56;  Gaps    6;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAE------EAKLKAKYPSLGQKPGG------- 50
           :::||   |: | |:|   |:|| |  :||      |  ::      |  ||
Db     386 EEDEERKQEDEGVERQPEAEEEGGLERKAENDNESFEDSIREPNTQYGSYPGDDENTQRD 445

Qy      51 -SDFLMKR-----------------------LQKGQKYFDSGDYNMAKAKMKNKQL-P 83
             | |::                        :|| :|        ::: : :||:: |
Db     446 VGDELVEESSKDKSPTPRSSTPNFNILSEESLDVQKDKK-------RVSRGRNQNKRVKP 498

Qy      84 SAGPDKNLVTGDHIPTPQ---------DLPQRKSS 109
           |   : || |   :|  :          |:| || :
Db     499 SVHAEDNLKTRKQVPRKRQKQVDSADVDVPTRKEA 533


RESULT  36
Q9VYU0
ID   Q9VYU0      PRELIMINARY;      PRT;   2274 AA.
AC   Q9VYU0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CG1905 PROTEIN.
GN   CG1905.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003487; AAF48098.1; -.
DR   FlyBase; FBgn0030337; CG1905.
SQ   SEQUENCE   2274 AA;  247009 MW;  B287311CBDF9BF20 CRC64;


  Query Match             13.7%;  Score 83;  DB 5;  Length 2274;
  Best Local Similarity   30.6%;  Pred. No. 55;
```

```
  Matches   38;  Conservative   15;  Mismatches   37;  Indels   34;  Gaps    7;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGIL--PERAE-------EAKLKAKYPSLGQKPGGS 51
           : :|: ||    || ||:|  :|:|  |   ||||       | ||| |   | :|
Db    1564 LKEKEREEKLKRET-EERQREKEREEKLKEKERAEKLKDLEKEVKLKEKEEQLKEK---- 1618

Qy      52 DFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPS---------AGPDKNLV--TGDHIPTP 100
           :  :|  :| :|          | | | | | |         : | : :|  |   :|
Db    1619 EKELKLKEKKEK---------DKVKEKEKSLESEKLLISATVSNPWRRVVEDTPPKLPAV 1669

Qy     101 QDLP 104
           || |
Db    1670 QDYP 1673


RESULT  37
Q9XGD3
ID   Q9XGD3      PRELIMINARY;      PRT;   795 AA.
AC   Q9XGD3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AT2G06430 PROTEIN.
GN   AT2G06430.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC007235; AAD26891.2; -.
DR   InterPro; IPR003653; SUMO_protease.
DR   Pfam; PF02902; Ulp1_C; 1.
SQ   SEQUENCE   795 AA;  91175 MW;  09E54A13ABF69B88 CRC64;


  Query Match             13.6%;  Score 82.5;  DB 10;  Length 795;
  Best Local Similarity   25.4%;  Pred. No. 19;
  Matches   33;  Conservative   22;  Mismatches   38;  Indels   37;  Gaps    6;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKR---LQK 60
           :::||   |: | |:|   |:|| |  :||:     | |:       : | :    :||
Db     405 EEDEERKQEDEGVERQPEAEEEGGLERKAED-----KSPTPRSTTPNFNILSEESLDVQK 459

Qy      61 GQKYFDSGDYNMAKAKMKNKQL-PSAGPDKNLVT---------------GDHIPTP---- 100
           :|        ::: : :||:: ||   : || |                 ||:
Db     460 DKK-------RVSRGRNENKRVKPSVHAEDNLKTRKQAQSKKRKIIGNDGDNADNDGDND 512

Qy     101 --QDLPQRKS 108
             |  |||||
Db     513 DFQPAPQRKS 522


RESULT  38
Q9VC00
ID   Q9VC00      PRELIMINARY;      PRT;  2768 AA.
AC   Q9VC00;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG13648 PROTEIN.
GN   CG13648.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003750; AAF56376.1; -.
DR   FlyBase; FBgn0039257; CG13648.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF00093; vwc; 1.
DR   SMART; SM00214; VWC; 4.
DR   PROSITE; PS01208; VWFC; 1.
SQ   SEQUENCE   2768 AA;  294032 MW;  CA929A21774E4684 CRC64;


  Query Match             13.6%;  Score 82;  DB 5;  Length 2768;
  Best Local Similarity   22.4%;  Pred. No. 86;
  Matches   32;  Conservative   24;  Mismatches   49;  Indels   38;  Gaps    5;

Qy       1 MSQKQEEENPAEETGEEK---QDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKR 57
           : : :|:: | ||: |:|   :|::|||  ||     :::: :     :|   :||
Db    1227 IDESEEDKKPVEESAEDKKPVEDSEEKEKPLPTVIPASEIEKESKPEDEKKTEADFAAPT 1286

Qy      58 LQ----------KGQKYFD----------SGDYNMAKAKMKNK----QLPSAGPDKNLV 92
           |            :|  |          ||:  :  |  | :    |:| |
```

```
Db   1287 EQPEATTPAQIADTAEKEVDDKLATTSAPVSGEDELKPADEKKRTETAQIPDA------- 1339

Qy     93 TGDHIPTPQDLPQRKSSLVTSKL 115
            ||   | |:  : | |  |
Db   1340 ---EIPASTDEPESSTELPTVDL 1359


RESULT  39
O64877
ID   O64877      PRELIMINARY;      PRT;   646 AA.
AC   O64877;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AT2G44430 PROTEIN.
GN   AT2G44430.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC004521; AAC16089.1; -.
DR   InterPro; IPR001487; Bromodomain.
DR   InterPro; IPR001005; Myb_DNA_bind.
DR   Pfam; PF00439; bromodomain; 1.
DR   Pfam; PF00249; myb_DNA-binding; 1.
DR   SMART; SM00297; BROMO; 1.
DR   SMART; SM00395; SANT; 1.
DR   PROSITE; PS50014; BROMODOMAIN_2; 1.
DR   PROSITE; PS50090; MYB_3; 1.
SQ   SEQUENCE   646 AA;  72405 MW;  104D7EB0E7E2AD42 CRC64;


  Query Match             13.5%;  Score 81.5;  DB 10;  Length 646;
  Best Local Similarity   25.8%;  Pred. No. 18;
  Matches   32;  Conservative   17;  Mismatches   52;  Indels   23;  Gaps    4;

Qy       3 QKQ-EEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKG 61
           ||| | :  :      ||||                | : |  |  :|   :||| |||:|
Db     515 QKQTESKTNSSNDNSSKQDT-------------GKTEKKTVSADKKKSVADFL-KRLKKN 560

Qy      62 QKYFDSGDYNM--------AKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTS 113
           :: | |          :| | :  :  | |  |  |    :       ||:|::  |:
Db     561 SPQKEAKDQNKSGGNVKKDSKTKPRELRSSSVGKKKAEVENTPVKRAPGRPQKKTAEATA 620

Qy     114 KLAG 117
           :|
Db     621 SASG 624


RESULT  40
Q9SI46
ID   Q9SI46      PRELIMINARY;      PRT;   789 AA.
AC   Q9SI46;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AT2G14130 PROTEIN.
GN   AT2G14130.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC007197; AAD25852.1; -.
DR   InterPro; IPR003653; SUMO_protease.
DR   Pfam; PF02902; Ulp1_C; 1.
SQ   SEQUENCE   789 AA;  90699 MW;  951560542417B8EA CRC64;


  Query Match             13.5%;  Score 81.5;  DB 10;  Length 789;
  Best Local Similarity   19.3%;  Pred. No. 23;
  Matches   28;  Conservative   28;  Mismatches   50;  Indels   39;  Gaps    4;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAE------EAKLKAKYPSLGQKPGGSDFLMK- 56
           :::||   |: | |:|   |:|| |  :||      |  ::      |  ||  :   :
Db     386 EEDEERKQEDEGVERQPEAEEEGGLERKAENDNESFEDSIREPNTQYGSYPGDDENTQRD 445

Qy      57 ----------------------RLQKGQKYFDSGDYNMAKAKMKNKQL-PSAGPDKNLVT 93
                                  : ::    : :::  :::||:: ||   : || |
Db     446 VGDEVAEESSKDKSHTPRSSTPNFNRIRRELAEEEMSLSFMQLQNKRVKPSVHAEDNLKT 505

Qy      94 GDHIPTPQ---------DLPQRKSS 109
           :|  :          |:| || :
Db     506 RKQVPRKRQKQVDSADVDVPTRKEA 530


Search completed: September  7, 2002, 10:50:54
Job time: 385 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:42:49 ; Search time 51.07 Seconds
(without alignments)
254.467 Million cell updates/sec

Title: US-09-824-178-3
Perfect score: 605
Sequence: 1 MSQKQEEENPAEETGEEKQD.....PTPQDLPQRKSSLVTSKLAG 117

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	605	100.0	117	19	AAW61551	Human endosulfine
2	605	100.0	121	19	AAW61552	Human endosulfine
3	605	100.0	121	22	AAB93514	Human protein sequ
4	595	98.3	122	21	AAG01489	Human secreted pro
5	492	81.3	150	21	AAB54082	Human pancreatic c
6	451	74.5	118	19	AAW60214	Human phosphoprote
7	427.5	70.7	125	22	AAO13512	Human polypeptide
8	402	66.4	96	19	AAW62655	Homo sapiens C4-2
9	296.5	49.0	82	21	AAG00770	Human secreted pro
10	296.5	49.0	82	21	AAG00772	Human secreted pro
11	288	47.6	101	21	AAG00771	Human secreted pro
12	283	46.8	66	21	AAG00769	Human secreted pro
13	226	37.4	119	22	ABB71486	Drosophila melanog
14	106	17.5	605	21	AAY57950	Human transmembran
15	104.5	17.3	200	18	AAW14545	Streptococcus pneu
16	100	16.5	605	20	AAY31741	Human podocalyxin-
17	100	16.5	605	22	AAU12251	Human PRO3742 poly
18	100	16.5	605	22	AAE03767	Human gene 4 encod
19	97.5	16.1	211	18	AAW14547	Streptococcus pneu
20	94.5	15.6	211	18	AAW14542	Streptococcus pneu
21	89	14.7	193	22	ABG11252	Novel human diagno
22	88	14.5	360	22	AAM93868	Human polypeptide,
23	87.5	14.5	1751	22	ABB69600	Drosophila melanog
24	85	14.0	615	22	ABG25869	Novel human diagno
25	85	14.0	659	22	ABG09291	Novel human diagno
26	84.5	14.0	131	22	ABG04385	Novel human diagno
27	84	13.9	147	22	ABG11269	Novel human diagno
28	84	13.9	969	22	ABB68123	Drosophila melanog
29	84	13.9	2286	22	AAB65635	Novel protein kina
30	83.5	13.8	281	22	ABB71057	Drosophila melanog
31	83.5	13.8	374	21	AAG16884	Arabidopsis thalia
32	83	13.7	619	15	AAR63437	Pneumococcal surfa
33	83	13.7	619	17	AAR87598	Pneumococcal surfa
34	83	13.7	619	17	AAR86911	Pneumococcal surfa
35	83	13.7	619	20	AAY41838	Streptococcus pneu
36	83	13.7	648	19	AAW70336	Pneumococcal surfa
37	83	13.7	648	19	AAW62274	Streptococcus pneu
38	83	13.7	648	20	AAY41837	Streptococcus pneu
39	83	13.7	648	20	AAW87879	A pneumococcal sur
40	83	13.7	653	20	AAW92456	S. pneumoniae trun
41	83	13.7	670	21	AAG51795	Arabidopsis thalia
42	83	13.7	684	16	AAR73912	Streptococcus pneu
43	83	13.7	688	21	AAG51794	Arabidopsis thalia
44	83	13.7	718	21	AAG42569	Arabidopsis thalia
45	83	13.7	722	21	AAG51793	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAW61551
ID AAW61551 standard; Protein; 117 AA.
XX
AC AAW61551;
XX
DT 19-OCT-1998 (first entry)
XX
DE Human endosulfine A protein.
XX
KW Endosulfine; central nervous system; pancreatic beta-cell; pain;
KW insulin; cytoprotective agent; diabetes; epilepsy; depression; ischemia.
XX
OS Homo sapiens.
XX
PN WO9830692-A2.
XX
PD 16-JUL-1998
XX
PF 07-JAN-1998; 98WO-US00137.
XX
PR 07-JAN-1997; 97US-0779775.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Anderson KL, Roch J, Scott VE, Sullivan JP;
XX
DR WPI; 1998-399140/34.
DR N-PSDB; AAV45210.
XX
PT Novel human endosulfine A and B - for screening for compounds
PT modulating endosulfine receptor activity, useful as cytoprotective
PT agent(s)
XX

```
PS   Claim 9; Page 58-59; 76pp; English.
XX
CC   The human endosulfine A and B genes encode proteins related to other
CC   members of the endosulfine family which have been shown to bind receptors
CC   from both the central nervous system and pancreatic beta-cells and to
CC   induce insulin secretion from a rodent beta-cell tumour line.  The
CC   protein can be used to screen for compounds that modulate endosulfine
CC   receptor activity, especially as a cytoprotective agent.
CC   Modulators of endosulfine activity may represent a treatment of
CC   diabetes, epilepsy, pain, depression and ischemia.
XX
SQ   Sequence   117 AA;

  Query Match             100.0%;  Score 605;  DB 19;  Length 117;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-58;
  Matches  117;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 msqkqeeenpaeetgeekqdtqekegilperaeeaklkakypslgqkpggsdflmkrlqk 60

Qy         61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 gqkyfdsgdynmakakmknkqlpsagpdknlvtgdhiptpqdlpqrksslvtsklag 117

RESULT   2
AAW61552
ID   AAW61552 standard; Protein; 121 AA.
XX
AC   AAW61552;
XX
DT   19-OCT-1998  (first entry)
XX
DE   Human endosulfine B protein.
XX
KW   Endosulfine; central nervous system; pancreatic beta-cell; pain;
KW   insulin; cytoprotective agent; diabetes; epilepsy; depression; ischemia.
XX
OS   Homo sapiens.
XX
PN   WO9830692-A2.
XX
PD   16-JUL-1998.
XX
PF   07-JAN-1998;   98WO-US00137.
XX
PR   07-JAN-1997;   97US-0779775.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Anderson KL,  Roch J,  Scott VE,  Sullivan JP;
XX
DR   WPI; 1998-399140/34.
DR   N-PSDB; AAV45211.
XX
PT   Novel human endosulfine A and B - for screening for compounds
PT   modulating endosulfine receptor activity, useful as cytoprotective
PT   agent(s)
XX
PS   Claim 9; Page 58; 76pp; English.
XX
CC   The human endosulfine A and B genes encode proteins related to other
CC   members of the endosulfine family which have been shown to bind receptors
CC   from both the central nervous system and pancreatic beta-cells and to
CC   induce insulin secretion from a rodent beta-cell tumour line.  The
CC   protein can be used to screen for compounds that modulate endosulfine
CC   receptor activity, especially as a cytoprotective agent.
CC   Modulators of endosulfine activity may represent a treatment of
CC   diabetes, epilepsy, pain, depression and ischemia.
XX
SQ   Sequence   121 AA;

  Query Match             100.0%;  Score 605;  DB 19;  Length 121;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-58;
  Matches  117;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 msqkqeeenpaeetgeekqdtqekegilperaeeaklkakypslgqkpggsdflmkrlqk 60

Qy         61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 gqkyfdsgdynmakakmknkqlpsagpdknlvtgdhiptpqdlpqrksslvtsklag 117

RESULT   3
AAB93514
ID   AAB93514 standard; Protein; 121 AA.
XX
AC   AAB93514;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:12846.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 12846; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
```

X bad date

X same

X bad date

```
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   121 AA;


  Query Match             100.0%;  Score 605;  DB 22;  Length 121;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-58;
  Matches  117;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 msqkqeeenpaeetgeekqdtqekegilperaeeaklkakypslgqkpggsdflmkrlqk 60

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 gqkyfdsgdynmakakmknkqlpsagpdknlvtgdhiptpqdlpqrksslvtsklag 117


RESULT   4
AAG01489
ID   AAG01489 standard; Protein; 122 AA.
XX
AC   AAG01489;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 5570.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC01495.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 5570; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   122 AA;
```

X not good date

```
  Query Match             98.3%;  Score 595;  DB 21;  Length 122;
  Best Local Similarity   98.3%;  Pred. No. 1.1e-56;
  Matches  115;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 msqkqeeenpaeetgeekqdtqekegilperaeeaklkakypslgqkpggsdflmkrlqk 60

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 gqkyfdsgdynmakakmknkqlpsagpdknlvtgdhipxpqdlpqrksslvxsklag 117


RESULT   5
AAB54082
ID   AAB54082 standard; Protein; 150 AA.
XX
AC   AAB54082;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human pancreatic cancer antigen protein sequence SEQ ID NO:534.
XX
KW   Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW   detection; diagnosis; identification; cytostatic; neuroprotective;
KW   nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW   antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW   linkage analysis; tissue identification; tissue typing; forensic;
KW   neural; immune system; muscular; reproductive; gastrointestinal;
KW   pulmonary; cardiovascular; renal; proliferative.
XX
OS   Homo sapiens.
XX
PN   WO200055320-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05989.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-579444/54.
DR   N-PSDB; AAC98847.
XX
PT   New nucleic acid that is a pancreatic cancer antigen for preventing,
PT   treating, or ameliorating a medical condition, particular pancreatic
PT   cancer, or for use in assays for diagnosing a pathological condition -
XX
PS   Claim 11; Page 972-973; 1379pp; English.
XX
CC   AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC   proteins, called pancreatic cancer antigens, given in AAB54008 to
CC   AAB54466. The human pancreatic cancer antigens have cytostatic,
CC   neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC   gynaecological, cardiant and antiinflammatory activities, and can be used
CC   in gene therapy. The polynucleotide and proteins can be used for
CC   preventing, treating, or ameliorating a medical condition or in assays
CC   for diagnosing a pathological condition or a susceptibility to one in a
CC   subject. Binding partners to the proteins and the activity of the
CC   proteins can be identified. The pancreatic cancer antigens can be used to
CC   detect, treat or prevent pancreatic disorders, especially cancer.
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
```

```
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   150 AA;


  Query Match             81.3%;  Score 492;  DB 21;  Length 150;
  Best Local Similarity   100.0%;  Pred. No. 2e-45;
  Matches   95;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      23 EKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQL 82
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56 ekegilperaeeaklkakypslgqkpggsdflmkrlqkgqkyfdsgdynmakakmknkql 115

Qy      83 PSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           ||||||||||||||||||||||||||||||||||
Db     116 psagpdknlvtgdhiptpqdlpqrksslvtsklag 150


RESULT   6
AAW60214
ID   AAW60214 standard; Protein; 118 AA.
XX
AC   AAW60214;
XX
DT   28-SEP-1998  (first entry)
XX
DE   Human phosphoprotein hPSHP.
XX
KW   hPSHP; phosphoprotein; neurotransmitter; signal transduction;
KW   human; Parkinson's disease; Huntington's diseases; diagnosis;
KW   drug screening.
XX
OS   Homo sapiens.
XX
PN   WO9818922-A1.
XX
PD   07-MAY-1998.
XX
PF   14-OCT-1997;   97WO-US18489.
XX
PR   28-OCT-1996;   96US-0739484.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Hillman JL;
XX
DR   WPI; 1998-272217/24.
DR   N-PSDB; AAV30285.
XX
PT   Human phosphoprotein polypeptide, hPSHP - useful for treating
PT   degenerative neuronal diseases e.g. Parkinson's and Huntington's,
PT   diagnosing hPSHP related diseases and drug screening
XX
PS   Claim 1; Page 48-49; 64pp; English.
XX
CC   This polypeptide comprises hPSHP, a novel human phosphoprotein that
CC   is believed to function in the signal transduction pathway of
CC   neurotransmitters in brain tissue and in cell development processes
CC   in peripheral organs, based on its chemical and structural homology
CC   with known phosphoproteins GI 162690 (71% identity) and GI 741603
CC   83% identity). Its amino acid sequence was deduced from an
CC   isolated genomic DNA fragment (see V302850. The invention also
CC   provides genetically engineered expression vectors, and host cells
CC   comprising nucleic acid sequences encoding hPSHP used in a claimed
CC   method for producing hPSHP, as well as probes and primers,
CC   antibodies, agonists and antagonists of hPSHP. hPSHP (or its
CC   fragments) is useful therapeutically to regenerate and enhance
CC   survival of nerve cells to halt degenerative processes in brain
CC   diseases such as Parkinson's and Huntington's disease. It can be
CC   used to induce antibodies to hPSHP, useful to diagnose conditions
CC   or diseases characterised by hPSHP expression, in assays to monitor
CC   patients receiving treatment, as agonists, antagonists or targeting
CC   molecules for pharmaceutical delivery, or to purify naturally
CC   occurring or recombinant hPSHP. The polypeptide or fragments are
CC   also useful in conventional drug screening. hPSHP agonists are
CC   useful e.g. for stimulating residual hPSHP in treatment of
CC   degenerative neuronal diseases, whilst antagonists are useful for
CC   inhibiting hPSHP, e.g. to regulate cell growth or suppress abnormal
CC   signal transduction in diseased tissue.
XX
SQ   Sequence   118 AA;


  Query Match             74.5%;  Score 451;  DB 19;  Length 118;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-41;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      31 RAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKN 90
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 raeeaklkakypslgqkpggsdflmkrlqkgqkyfdsgdynmakakmknkqlpsagpdkn 87

Qy      91 LVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           ||||||||||||||||||||||||||
Db      88 lvtgdhiptpqdlpqrksslvtsklag 114


RESULT   7
AAO13512
ID   AAO13512 standard; Protein; 125 AA.
XX
AC   AAO13512;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 27404.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI93443.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 20; SEQ ID NO 27404; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
```

```
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   125 AA;


  Query Match             70.7%;  Score 427.5;  DB 22;  Length 125;
  Best Local Similarity   76.4%;  Pred. No. 1.5e-38;
  Matches   84;  Conservative    9;  Mismatches   16;  Indels    1;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDS 67
           | |   : ||::: ::|    ||:|||||||||:|| |||||||||| ||||||||||||
Db      17 evpeaasaeeqkemedkv-tspekaeeaklkaryphlgqkpggsdflrkrlqkgqkyfds 75

Qy      68 GDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           |||||||||||||||:| |||  |||||||||||||||| ||| |||||
Db      76 gdynmakakmknkqlptaapdktevtgdhiptpqdlpqrkpslvasklag 125


RESULT   8
AAW62655
ID   AAW62655 standard; Protein; 96 AA.
XX
AC   AAW62655;
XX
DT   27-OCT-1998  (first entry)
XX
DE   Homo sapiens C4-2 tumour suppressor protein.
XX
KW   C4-2; tumour suppressor gene; cancer; treatment; glioblastoma;
KW   cell proliferative disorders; brain tumour.
XX
OS   Homo sapiens.
XX
PN   WO9820021-A1.
XX
PD   14-MAY-1998.
XX
PF   07-NOV-1997;   97WO-US20408.
XX
PR   08-NOV-1996;   96US-0744905.
XX
PA   (NWBI-) NORTHWEST BIOTHERAPEUTICS LLC.
XX
PI   Boynton AL,  Murphy GP,  Sehgal A;
XX
DR   WPI; 1998-286852/25.
DR   N-PSDB; AAV38840.
XX
PT   Tumour suppressor, C4-2 - useful for treatment of proliferative
PT   diseases e.g. cancer
XX
PS   Example; Fig 8; 112pp; English.
XX
CC   The sequence is that encoded by the C4-2 tumour suppressor gene.
CC   It can be used in the treatment of cell proliferative diseases,
CC   e.g. cancer, in particular glioblastomas and other malignant
CC   brain tumours.
XX
SQ   Sequence   96 AA;


  Query Match             66.4%;  Score 402;  DB 19;  Length 96;
  Best Local Similarity   86.5%;  Pred. No. 6.4e-36;
  Matches   77;  Conservative    4;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      29 PERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPD 88
           ||:|||||||:|| |||||||:|| |||||||||||||||||||||||||||:| ||
Db       8 pekaeeaklkaryphlgqkpggsnflrkrlqkgqkyfdsgdynmakakmknkqlptaapd 67

Qy      89 KNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           |  |||||||||||||| | ||| |||||
Db      68 ktevtgdhiptpqdlpqlkpslvasklag 96


RESULT   9
AAG00770
ID   AAG00770 standard; Protein; 82 AA.
XX
AC   AAG00770;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 4851.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC00776.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 4851; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   82 AA;


  Query Match             49.0%;  Score 296.5;  DB 21;  Length 82;
  Best Local Similarity   72.5%;  Pred. No. 1.4e-24;
  Matches   58;  Conservative    9;  Mismatches   12;  Indels    1;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDS 67
           | |   : ||::: ::|    ||:|||||||:|| |||||||||| ||||||||||||
Db       4 evpeaasaeeqkemedkv-tspekaeeaklkaryphlgqkpggsdflrkrlqkgqkyfds 62

Qy      68 GDYNMAKAKMKNKQLPSAGP 87
           ||||||||||||||||:| |
Db      63 gdynmakakmknkqlptaap 82


RESULT  10
```

```
AAG00772
ID   AAG00772 standard; Protein; 82 AA.
XX
AC   AAG00772;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 4853.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC00778.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 4853; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   82 AA;


  Query Match             49.0%;  Score 296.5;  DB 21;  Length 82;
  Best Local Similarity   72.5%;  Pred. No. 1.4e-24;
  Matches   58;  Conservative    9;  Mismatches   12;  Indels    1;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDS 67
           | |   : ||::: ::|    ||:||||||||:|| |||||||||| ||||||||||
Db       4 evpeaasaeeqkemedkv-tspekaeeaklkaryphlgqkpggsdflrkrlqkgqkyfds 62

Qy      68 GDYNMAKAKMKNKQLPSAGP 87
           ||||||||||||||||:| |
Db      63 gdynmakakmknkqlptaap 82


RESULT  11
AAG00771
ID   AAG00771 standard; Protein; 101 AA.
XX
AC   AAG00771;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 4852.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC00777.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 4852; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   101 AA;


  Query Match             47.6%;  Score 288;  DB 21;  Length 101;
  Best Local Similarity   60.4%;  Pred. No. 1.5e-23;
  Matches   61;  Conservative    6;  Mismatches   20;  Indels   14;  Gaps    1;

Qy       1 MSQKQEEENPAEETGEEKQ----------DTQEKEGILPERAEEAKLKAKYPSLGQ 46
           || :  |   |||   |               :  | :   ||:||||||||:|| |||
Db       1 msaevpeaasaeeqksehnmlpwslqpsipnsleemedkvtspekaeeaklkaryphlgq 60

Qy      47 KPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGP 87
           |||||||| ||||||||||||||||||||||||||||:| |
Db      61 kpggsdflrkrlqkgqkyfdsgdynmakakmknkqlptaap 101


RESULT  12
AAG00769
ID   AAG00769 standard; Protein; 66 AA.
XX
AC   AAG00769;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 4850.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
```

```
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC00775.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 4850; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   66 AA;


  Query Match             46.8%;  Score 283;  DB 21;  Length 66;
  Best Local Similarity   89.8%;  Pred. No. 3e-23;
  Matches   53;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      29 PERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGP 87
           ||:|||||||:|| |||||||||| |||||||||||||||||||||||||||:| |
Db       8 pekaeeaklkaryphlgqkpggsdflrkrlqkgqkyfdsgdynmakakmknkqlptaap 66


RESULT  13
ABB71486
ID   ABB71486 standard; Protein; 119 AA.
XX
AC   ABB71486;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 41250.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL15589.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 41250; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   119 AA;


  Query Match             37.4%;  Score 226;  DB 22;  Length 119;
  Best Local Similarity   43.7%;  Pred. No. 9.3e-17;
  Matches   52;  Conservative   17;  Mismatches   32;  Indels   18;  Gaps    3;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGG-SDFLMKRLQ 59
           ||  :|  |    | :: : |::      |: || |||:||||  : ||| | || ||||
Db       1 mssaeensnspattpqdtetteqanltdlekieeeeklkskypsgmrvpgghsaflqkrlq 60

Qy      60 KGQKYFDSGDYNMAKAK-------MKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLV 111
           ||||:|||||| ||| |         ||         : ||: |||| : |  ||:|::
Db      61 kgqkffdsgdyqmakqkgggvkqvfank----------vttgeaiptpetvparktsii 109


RESULT  14
AAY57950
ID   AAY57950 standard; Protein; 605 AA.
XX
AC   AAY57950;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Human transmembrane protein HTMPN-74.
XX
KW   Human; transmembrane protein; HTMPN; diagnosis; immunospecific;
KW   antiproliferative; neuroprotective; immune disorder;
KW   reproductive disorder; smooth muscle disorder; neurological disorder;
KW   gastrointestinal disorder; developmental disorder;
KW   cell proliferative disorder.
XX
OS   Homo sapiens.
XX
PN   WO9961471-A2.
XX
PD   02-DEC-1999.
XX
PF   28-MAY-1999;   99WO-US11904.
XX
PR   29-MAY-1998;   98US-0087260.
PR   02-JUL-1998;   98US-0091674.
PR   02-OCT-1998;   98US-0102954.
PR   24-NOV-1998;   98US-0109869.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tang YT,  Lal P,  Hillman JL,  Yue H,  Guegler KJ,  Corley NC;
PI   Bandman O,  Patterson C,  Gorgone GA,  Kaser MR,  Baughn MR;
PI   Au-Young J;
```

```
XX
DR   WPI; 2000-072605/06.
DR   N-PSDB; AAZ56771.
XX
PT   Proteins, polynucleotides, vectors, host cells and antibodies used to
PT   diagnose, treat or prevent immune, reproductive, smooth muscle,
PT   neurological, gastrointestinal, developmental and cell proliferative
PT   disorders -
XX
PS   Claim 1; Page 172-174; 229pp; English.
XX
CC   AAZ56698 to AAZ56776 encode AAY57877 to AAY57955 which represent human
CC   transmembrane proteins designated HTMPN-1 to HTMPN-79, respectively.
CC   The transmembrane protein have immunospecific, antiproliferative and
CC   neuroprotective activities. The human transmembrane proteins,
CC   polynucleotides encoding them and other compositions and methods from
CC   the present invention, can be used for the diagnosis, treatment or
CC   prevention of immune, reproductive, smooth muscle, neurological,
CC   gastrointestinal, developmental and cell proliferative disorders. The
CC   HTMPN's can be used to treat or prevent disorders associated with a
CC   decreased expression or activity of HTMPN.
XX
SQ   Sequence   605 AA;


  Query Match             17.5%;  Score 106;  DB 21;  Length 605;
  Best Local Similarity   28.9%;  Pred. No. 0.0069;
  Matches   41;  Conservative   20;  Mismatches   43;  Indels   38;  Gaps    7;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERA--EEAKLKAKYPSL---GQKPGGSDFLMKR 57
           :::|||   |:   |||| :|:| :||     |||| : :  ||    | || :
Db     165 eeeeeeeemekeevekqdveeeeellpvngsqeeakpqvrdfsltsssqtpgatk----- 219

Qy      58 LQKGQKYFDSGDYNMAKAKMKNKQ-----LPSAGPDKNLVT-GDHIPTPQ---------- 101
             :: ||||   :  ::::       |||  |   :|| ||   | |
Db     220 ----srhedsgdqassgvevessmgpslllpsvtp--tivtpgdqdstsqeaeatvlpaa 273

Qy     102 ------DLPQRKSSLVTSKLAG 117
                : ||  |   |:  ||
Db     274 glgvefeapqeaseeatagaag 295


RESULT  15
AAW14545
ID   AAW14545 standard; Protein; 200 AA.
XX
AC   AAW14545;
XX
DT   28-OCT-1997  (first entry)
XX
DE   Streptococcus pneumoniae PspA N-terminal end.
XX
KW   PspA; pneumococcal surface protein; vaccine; otitis media;
KW   meningitis; bacteraemia; pneumonia.
XX
OS   Streptococcus pneumoniae strain Db16aapro.
XX
PN   WO9709994-A1.
XX
PD   20-MAR-1997.
XX
PF   16-SEP-1996;   96WO-US14819.
XX
PR   15-SEP-1995;   95US-0529055.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  Brooks-Walter A,  Crain MJ,  Hollingshead S;
PI   McDaniel LS, Swiatlo E,  Tart R,  Yother J;
XX
DR   WPI; 1997-202002/18.
XX
PT   Streptococcus pneumoniae surface protein PspC and truncated PspA -
PT   used in vaccines for protecting animals against S.pneumoniae
PT   infection
XX
PS   Example 6; Fig 13; 296pp; English.
XX
CC   This sequence shows the N-terminal end of pneumococcal surface
CC   protein A (PspA) of Streptococcus pneumoniae strain Db16aapro.
CC   Comparison of the N-terminal and central regions (AAW14533-57 and
CC   AAW14562-91) of PspA polypeptides from different pneumococcal strains
CC   can be used to divide the strains into several families based on
CC   sequence homologies.  PspA polypeptides, or fragments of them, can
CC   be used in vaccines to protect animals against S. pneumoniae
CC   infection and hence for the prevention of diseases such as otitis
CC   media, meningitis, bacteraemia and pneumonia.  The sequence of the
CC   3' half of the PspA alpha-helical region and the immediate 5' tip
CC   of the coding sequence are likely to be the critical sequences for
CC   predicting PspA cross-reactions and vaccine composition.
XX
SQ   Sequence   200 AA;


  Query Match             17.3%;  Score 104.5;  DB 18;  Length 200;
  Best Local Similarity   28.8%;  Pred. No. 0.0025;
  Matches   36;  Conservative   26;  Mismatches   44;  Indels   19;  Gaps    6;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           |:: || |::  || :  |||      | ||   |  : || |:    |: :|     :|
Db      50 krdaen-akkaleeakraqekyadyqrrieekaakethasleqqeankdyqlkl----kk 104

Qy      64 YFDSGDYN--------MAKAKMKNKQLPSAGPDKNLVTGD-HIPTPQDL--PQRKSSLVT 112
           | |  : :         | :|: |:|: |:   : | :  :  :| ||:|   :||| :
Db     105 yldgrnlsnssvlkkemeeaekkdkekpa---efnkirreivvpnpqelemarrksevak 161

Qy     113 SKLAG 117
           :| :|
Db     162 tkesg 166


RESULT  16
AAY31741
ID   AAY31741 standard; Protein; 605 AA.
XX
AC   AAY31741;
XX
DT   22-NOV-1999  (first entry)
XX
DE   Human podocalyxin-like sialomucin PCLP-2.
XX
KW   PCLP-2; podocalyxin-like protein; selectin; ligand; human;
KW   inflammation; autoimmune disease; tissue rejection; therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..32
FT                   /note= "signal peptide"
FT   Protein         33..605
FT                   /note= "mature protein"
XX
PN   WO9941363-A1.
XX
PD   19-AUG-1999.
XX
PF   28-JAN-1999;   99WO-US01780.
XX
PR   10-DEC-1998;   98US-0111663.
PR   11-FEB-1998;   98US-0074389.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Rosen SD,  Sassetti CM;
XX
DR   WPI; 1999-540131/45.
```

```
DR   N-PSDB; AAX87926.
XX
PT   Podocalyxin-like protein, e.g. useful for diagnosing and treating
PT   diseases associated with selectin mediated binding events
XX
PS   Claim 7; Fig 1; 63pp; English.
XX
CC   The present sequence represents a new podocalyxin-like protein
CC   (PCLP-2) having selectin ligand activity.  PCLP-2 is a sialomucin
CC   having 4 distinct regions: a mucin domain, a cysteine-rich domain,
CC   a transmembrane domain; and a cytoplasmic domain.  Its N-terminal
CC   domain differentiates it from PCLP.  Homology is shown to CD34 in
CC   the cyoplasmic tail domain.  PCLP-2 polypeptides and nucleic acids
CC   are useful for diagnosing and treating (via gene therapy) diseases
CC   associated with selectin mediated binding events, including acute
CC   and chronic inflammation, autoimmune diseases and tissue rejection,
CC   and are useful in assays for screening for therapeutic agents.
XX
SQ   Sequence   605 AA;


  Query Match             16.5%;  Score 100;  DB 20;  Length 605;
  Best Local Similarity   27.0%;  Pred. No. 0.031;
  Matches   38;  Conservative   21;  Mismatches   46;  Indels   36;  Gaps    6;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERA--EEAKLKAKYPSL---GQKPGGSDFLMKR 57
           :::|||   |:   |||: :|:| :||     |||| : :  ||    | || :
Db     165 eeeeeeeerekeevekqeeeeeeellpvngsqeeakpqvrdfsltsssqtpgatk----- 219

Qy      58 LQKGQKYFDSGDYNMAKAKMKNKQ-----LPSAGPDKNLVTGDHIPTPQ---------- 101
             :: ||||   :  ::::       ||| |   :  ||   | |
Db     220 ----srhedsgdqassgvevessmgpslllpsvtp-ttvtpgdqdstsqeaeatvlpaag 274

Qy     102 -----DLPQRKSSLVTSKLAG 117
               : ||  |   |:  ||
Db     275 lgvefeapqeaseeatagaag 295


RESULT  17
AAU12251
ID   AAU12251 standard; Protein; 605 AA.
XX
AC   AAU12251;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human PRO3742 polypeptide sequence.
XX
KW   Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW   breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW   cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW   adipocyte; A-peptide; factor VIIA; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200140466-A2.
XX
PD   07-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-US32678.
XX
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28564.
PR   02-DEC-1999;   99WO-US28565.
PR   09-DEC-1999;   99US-0170262.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   20-DEC-1999;   99WO-US30999.
PR   30-DEC-1999;   99WO-US31243.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   01-MAR-2000; 2000WO-US05601.
PR   20-MAR-2000; 2000WO-US07377.
PR   21-MAR-2000; 2000WO-US07532.
PR   30-MAR-2000; 2000WO-US08439.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   10-NOV-2000; 2000WO-US30873.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Beresini M,  Deforge L,  Desnoyers L,  Filvaroff E,  Gao W;
PI   Gerritsen ME,  Goddard A,  Godowski PJ,  Gurney AL,  Sherwood S;
PI   Smith V,  Stewart TA,  Tumas D,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2001-408281/43.
DR   N-PSDB; AAS21323.
XX
PT   Isolated, secretory and transmembrane PRO polypeptide used to detect
PT   other PRO polypeptides, link bioactive molecules to cells expressing
PT   PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT   lung, breast, prostate, cervical  -
XX
PS   Claim 12; Fig 160; 813pp; English.
XX
CC   AAU12172-AAU12446 represent novel human secretory and transmembrane
CC   PRO polypeptides. The PRO polypeptides are useful to detect other
CC   PRO polypeptides, to link bioactive molecules to cells expressing
CC   PRO polypeptides, to modulate biological activities of cells expressing
CC   PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC   breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC   polypeptide expression in a cell sample to that in a control sample.
CC   Some of the 275 sequences are also useful to stimulate the release of
CC   tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC   proliferation or differentiation of chondrocytes, the proliferation or
CC   gene expression in pericyte cells, the release of proteoglycans from
CC   cartilage, the proliferation of inner ear utricular supporting cells or
CC   of T-lymphocytes, the release of a cytokine from peripheral blood
CC   monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC   the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC   skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
CC   to factor VIIA. The PRO polypeptides can be used in assays to identify
CC   molecules involved in binding interactions. The polynucleotides encoding
CC   PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC   transgenic or knock out animals and can be used in gene therapy.
XX
SQ   Sequence   605 AA;


  Query Match             16.5%;  Score 100;  DB 22;  Length 605;
  Best Local Similarity   27.0%;  Pred. No. 0.031;
  Matches   38;  Conservative   21;  Mismatches   46;  Indels   36;  Gaps    6;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERA--EEAKLKAKYPSL---GQKPGGSDFLMKR 57
           :::|||   |:   |||: :|:| :||     |||| : :  ||    | || :
Db     165 eeeeeeeerekeevekqeeeeeeellpvngsqeeakpqvrdfsltsssqtpgatk----- 219

Qy      58 LQKGQKYFDSGDYNMAKAKMKNKQ-----LPSAGPDKNLVTGDHIPTPQ---------- 101
             :: ||||   :  ::::       ||| |   :  ||   | |
Db     220 ----srhedsgdqassgvevessmgpslllpsvtp-ttvtpgdqdstsqeaeatvlpaag 274

Qy     102 -----DLPQRKSSLVTSKLAG 117
               : ||  |   |:  ||
Db     275 lgvefeapqeaseeatagaag 295
```

```
RESULT  18
AAE03767
ID   AAE03767 standard; Protein; 605 AA.
XX
AC   AAE03767;
XX
DT   07-AUG-2001  (first entry)
XX
DE   Human gene 4 encoded secreted protein HISEL50, SEQ ID NO:37.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive;
KW   binding partner identification.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..33
FT                   /label= signal_peptide
FT   Protein         34..605
FT                   /note= "Mature secreted protein"
XX
PN   WO200132837-A1.
XX
PD   10-MAY-2001.
XX
PF   17-OCT-2000; 2000WO-US28664.
XX
PR   02-NOV-1999;   99US-0163085.
PR   17-DEC-1999;   99US-0172411.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Fiscella M,  Wei P,  Lafleur DW,  Olsen HS,  Baker K,  Ebner R;
PI   Komatsoulis G,  Rosen CA,  Ruben SM,  Duan RD,  Young PE,  Florence KA;
PI   Moore PA,  Birse CE,  Ni J,  Soppet DR,  Shi Y;
XX
DR   WPI; 2001-328782/34.
DR   N-PSDB; AAD08194.
XX
PT   Novel human secreted proteins and nucleic acids for diagnosing,
PT   preventing and treating neurological, cardiovascular, infectious,
PT   autoimmune, gastrointestinal, bone disorders, cancer, particularly
PT   ovarian cancer  -
XX
PS   Claim 11; Page 390-392; 421pp; English.
XX
CC   AAD08191-AAD08213 represent cDNAs corresponding to 19 human secreted
CC   protein genes and AAE03764-AAE03786 represent the proteins they encode.
CC   AAE03787-AAE03800 represent human secreted protein fragments or variants.
CC   The genes and their secreted proteins are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   19 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin aging due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein of the invention.
XX
SQ   Sequence   605 AA;


  Query Match           16.5%;  Score 100;  DB 22;  Length 605;
  Best Local Similarity   27.0%;  Pred. No. 0.031;
  Matches   38;  Conservative   21;  Mismatches   46;  Indels   36;  Gaps    6;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERA--EEAKLKAKYPSL---GQKPGGSDFLMKR 57
           :::|||   |:   |||: :|:| :||     ||||  : :  ||    | || :
Db     165 eeeeeeeerekeevekqeeeeeeellpvngsqeeakpqvrdfsltsssqtpgatk----- 219

Qy      58 LQKGQKYFDSGDYNMAKAKMKNKQ-----LPSAGPDKNLVTGDHIPTPQ---------- 101
             :: ||||   :  ::::       |||  |   :  ||   | |
Db     220 ----srhedsgdqassgvevessmgpslllpsvtp-ttvtpgdqdstsqeaeatvlpaag 274

Qy     102 -----DLPQRKSSLVTSKLAG 117
                : ||  |   |:  ||
Db     275 lgvefeapqeaseeatagaag 295


RESULT  19
AAW14547
ID   AAW14547 standard; Protein; 211 AA.
XX
AC   AAW14547;
XX
DT   28-OCT-1997  (first entry)
XX
DE   Streptococcus pneumoniae PspA N-terminal end.
XX
KW   PspA; pneumococcal surface protein; vaccine; otitis media;
KW   meningitis; bacteraemia; pneumonia.
XX
OS   Streptococcus pneumoniae strain Ef10197pro.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 209
FT                   /note= "unidentified amino acid"
XX
PN   WO9709994-A1.
XX
PD   20-MAR-1997.
XX
PF   16-SEP-1996;   96WO-US14819.
XX
PR   15-SEP-1995;   95US-0529055.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  Brooks-Walter A,  Crain MJ,  Hollingshead S;
PI   McDaniel LS, Swiatlo E,  Tart R,  Yother J;
XX
DR   WPI; 1997-202002/18.
XX
PT   Streptococcus pneumoniae surface protein PspC and truncated PspA -
PT   used in vaccines for protecting animals against S.pneumoniae
PT   infection
XX
PS   Example 6; Fig 13; 296pp; English.
XX
CC   This sequence shows the N-terminal end of pneumococcal surface
```

```
CC   protein A (PspA) of Streptococcus pneumoniae strain Ef10197pro.
CC   Comparison of the N-terminal and central regions (AAW14533-57 and
CC   AAW14562-91) of PspA polypeptides from different pneumococcal strains
CC   can be used to divide the strains into several families based on
CC   sequence homologies.  PspA polypeptides, or fragments of them, can
CC   be used in vaccines to protect animals against S. pneumoniae
CC   infection and hence for the prevention of diseases such as otitis
CC   media, meningitis, bacteraemia and pneumonia.  The sequence of the
CC   3' half of the PspA alpha-helical region and the immediate 5' tip
CC   of the coding sequence are likely to be the critical sequences for
CC   predicting PspA cross-reactions and vaccine composition.
XX
SQ   Sequence   211 AA;


  Query Match             16.1%;  Score 97.5;  DB 18;  Length 211;
  Best Local Similarity   29.4%;  Pred. No. 0.015;
  Matches   37;  Conservative   20;  Mismatches   48;  Indels   21;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           |:: || |::  || :  |||      | ||   | :  || |:    |: :|     :|
Db      51 krdaen-akkaleeakraqekyadyqrrieekaakeqqasleqqeamkdyqlkl----kk 105

Qy      64 YFD----------SGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDL--PQRKSSLV 111
           | |            :   |: | | ||       : :|    :| ||:|   :||| :|
Db     106 yldgrnlsnssvlkkemeeaekkdkekqaefnkirreiv----vpnpqelemarrksevv 161

Qy     112 TSKLAG 117
             :| :|
Db     162 kakesg 167


RESULT  20
AAW14542
ID   AAW14542 standard; Protein; 211 AA.
XX
AC   AAW14542;
XX
DT   28-OCT-1997  (first entry)
XX
DE   Streptococcus pneumoniae PspA N-terminal end.
XX
KW   PspA; pneumococcal surface protein; vaccine; otitis media;
KW   meningitis; bacteraemia; pneumonia.
XX
OS   Streptococcus pneumoniae strain Bg9739pro.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 197
FT                   /note= "unidentified amino acid"
FT   Misc-difference 198
FT                   /note= "unidentified amino acid"
XX
PN   WO9709994-A1.
XX
PD   20-MAR-1997.
XX
PF   16-SEP-1996;   96WO-US14819.
XX
PR   15-SEP-1995;   95US-0529055.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  Brooks-Walter A,  Crain MJ,  Hollingshead S;
PI   McDaniel LS, Swiatlo E,  Tart R,  Yother J;
XX
DR   WPI; 1997-202002/18.
XX
PT   Streptococcus pneumoniae surface protein PspC and truncated PspA -
PT   used in vaccines for protecting animals against S.pneumoniae
PT   infection
XX
PS   Example 6; Fig 13; 296pp; English.
XX
CC   This sequence shows the N-terminal end of pneumococcal surface
CC   protein A (PspA) of Streptococcus pneumoniae strain Bg9739pro.
CC   Comparison of the N-terminal and central regions (AAW14533-57 and
CC   AAW14562-91) of PspA polypeptides from different pneumococcal strains
CC   can be used to divide the strains into several families based on
CC   sequence homologies.  PspA polypeptides, or fragments of them, can
CC   be used in vaccines to protect animals against S. pneumoniae
CC   infection and hence for the prevention of diseases such as otitis
CC   media, meningitis, bacteraemia and pneumonia.  The sequence of the
CC   3' half of the PspA alpha-helical region and the immediate 5' tip
CC   of the coding sequence are likely to be the critical sequences for
CC   predicting PspA cross-reactions and vaccine composition.
XX
SQ   Sequence   211 AA;


  Query Match             15.6%;  Score 94.5;  DB 18;  Length 211;
  Best Local Similarity   28.0%;  Pred. No. 0.032;
  Matches   35;  Conservative   26;  Mismatches   45;  Indels   19;  Gaps    6;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           |:: || |::  || :  |||      | ||   |    || |:    |: :|     :|
Db      50 krdaen-akkaleeakraqekyadyqrrieekaaketqasleqqeankdyqlkl----kk 104

Qy      64 YFDSGDYN--------MAKAKMKNKQLPSAGPDKNLVTGD-HIPTPQDL--PQRKSSLVT 112
           | |  : :         | :|: |:|:  :   ::| :  :  :| ||:|   :||| :|
Db     105 yldgrnlsnssvlkkemeeaekkdkenqa---eenkirreivvpnpqelemarrksevvk 161

Qy     113 SKLAG 117
           :  :|
Db     162 atesg 166


RESULT  21
ABG11252
ID   ABG11252 standard; Protein; 193 AA.
XX
AC   ABG11252;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #11243.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS75439.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 41611; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
```

```
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   193 AA;


  Query Match             14.7%;  Score 89;  DB 22;  Length 193;
  Best Local Similarity   27.3%;  Pred. No. 0.11;
  Matches   30;  Conservative   17;  Mismatches   41;  Indels   22;  Gaps    4;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           |:||:|||   :|  ||::: :|:|    |  || | | |              :: ::
Db       1 maqkkeeeeekkeeeeeeeeeeeee--eeeeeeekkkkk-------------eekrer 44

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSL 110
             :||      |      :|| | |  |     | :| |        |:|:
Db      45 skkykqqqqqn----tkrnknlts--pqdetqtanhkpeehpeentknsv 88


RESULT  22
AAM93868
ID   AAM93868 standard; Protein; 360 AA.
XX
AC   AAM93868;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide, SEQ ID NO: 3973.
XX
KW   Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS   Homo sapiens.
XX
PN   EP1130094-A2.
XX
PD   05-SEP-2001.
XX
PF   07-JUL-2000; 2000EP-0114089.
XX
PR   08-JUL-1999;   99JP-0194486.
PR   11-JAN-2000; 2000JP-0118774.
PR   02-MAY-2000; 2000JP-0183765.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI   Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR   WPI; 2001-524255/58.
DR   N-PSDB; AAK94826.
XX
PT   830 Primers useful for synthesizing full length cDNA clones and their
PT   use in genetic manipulation -
XX
PS   Claim 8; SEQ ID NO 3973; 1380pp + sequence listing; English.
XX
CC   The invention relates to primers for synthesising full length cDNA
CC   clones. 830 cDNA molecules encoding a human protein have been
CC   isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC   molecules have been determined. Primers for synthesising the full length
CC   cDNA are useful for clarifying the function of the protein encoded by
CC   the cDNA. The full length clones were obtained by construction of full
CC   length enriched cDNA libraries that were synthesised by the oligo-capping
CC   method. The primers enable the production of the full length cDNA easily
CC   without any special methods. The present sequence is a polypeptide
CC   encoded by a full length human cDNA of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ   Sequence   360 AA;


  Query Match             14.5%;  Score 88;  DB 22;  Length 360;
  Best Local Similarity   25.9%;  Pred. No. 0.32;
  Matches   37;  Conservative   22;  Mismatches   50;  Indels   34;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPER-----------AEEAKLKAKYP----SLGQK 47
           ||: :   ::  |||::  :|:| | |:           :|:|:|:       :  ||
Db      89 kqqnqeitdqleeekkechqkqgalreqlqvhiqtigilvsekaelqtalahtqhaarqk 148

Qy      48 PGGSDFLMKRLQKGQKYFDSGDYNMA---------KAKMKNKQLPSAGPDKNLVTGDHIP 98
             | |: |  |||  ::    |:   |         ||   ||:|        |    :
Db     149 egesedlasrlqysrrrv--geleralsavstqqkkadrynkeltkerdalrlelykntq 206

Qy      99 TPQDLPQRKSS-------LVTSK 114
           : :|| | ||         ||| |
Db     207 snedlkqekseleeklrvlvtek 229


RESULT  23
ABB69600
ID   ABB69600 standard; Protein; 1751 AA.
XX
AC   ABB69600;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 35592.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL13703.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 35592; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
```

```
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1751 AA;


  Query Match             14.5%;  Score 87.5;  DB 22;  Length 1751;
  Best Local Similarity   26.3%;  Pred. No. 2.6;
  Matches   35;  Conservative   18;  Mismatches   57;  Indels   23;  Gaps    4;

Qy       3 QKQEEENPA-------EETGEEKQDTQEKEGILPERAEEAKLKAKYPSLG---QKPGGSD 52
           |  :| |||         :|||::   :| |   :|: :   :   |: |    :
Db     976 qssdesnparieeendsnsgeereg--kkrgtsgqrstdrsaerispafgfnrlelrskq 1033

Qy      53 FLMKRLQKGQKYFDSGDYNMAKAK----------MKNKQLPSAGPDKNLVTGDHIPTPQ 101
           |  | :  |  :  ||   |  ||            :|:||    ||   || |  |  :
Db    1034 fkskpipigahhtvsgheamdlakrlneelashqqqqnqqlrpispdiravtpdceprsr 1093

Qy     102 DLPQRKSSLVTSK 114
             || :| | :|
Db    1094 sfeqrptsgvcak 1106


RESULT  24
ABG25869
ID   ABG25869 standard; Protein; 615 AA.
XX
AC   ABG25869;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #25860.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS90056.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 56228; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   615 AA;


  Query Match             14.0%;  Score 85;  DB 22;  Length 615;
  Best Local Similarity   25.0%;  Pred. No. 1.3;
  Matches   29;  Conservative   23;  Mismatches   36;  Indels   28;  Gaps    4;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQ 62
           ||||||   ||  ||::: :|:|    |  || :        |:: ||         :|:
Db     361 qkqeeeeeeeeeeeeeeeeeeeeeeeeeeege--------geregg---------ege 403

Qy      63 KYFDSGDY-----NMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTS 113
           :  : |::     :  :|| :  :   : ||      : ||    :  |   ||::
Db     404 regeegehdrfsthkpqakdetmkyirskpd------nriperhyiagrprelvst 453


RESULT  25
ABG09291
ID   ABG09291 standard; Protein; 659 AA.
XX
AC   ABG09291;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #9282.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS73478.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 39650; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
```

```
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   659 AA;


  Query Match             14.0%;  Score 85;  DB 22;  Length 659;
  Best Local Similarity   25.0%;  Pred. No. 1.4;
  Matches   29;  Conservative   23;  Mismatches   36;  Indels   28;  Gaps    4;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQ 62
           ||||||   ||  ||::: :|:|    |  || :        |:: ||         :|:
Db     405 qkqeeeeeeeeeeeeeeeeeeeeeeeeeeeege--------geregg---------ege 447

Qy      63 KYFDSGDY-----NMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTS 113
           :  : |::     :  :|| :  :   : ||      : ||    :  |   ||::
Db     448 regeegehdrfsthkpqakdetmkyirskpd------nriperhyiagrprelvst 497


RESULT  26
ABG04385
ID   ABG04385 standard; Protein; 131 AA.
XX
AC   ABG04385;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #4376.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS68572.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 34744; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   131 AA;


  Query Match             14.0%;  Score 84.5;  DB 22;  Length 131;
  Best Local Similarity   26.9%;  Pred. No. 0.21;
  Matches   28;  Conservative   17;  Mismatches   52;  Indels    7;  Gaps    2;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGS------DFLMK 56
           :::|||   ||  ||::: :|:|    |  || : : :   ||:  |||       |
Db      28 eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeegeeeelgrlhggsgkvrglgften 87

Qy      57 RLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTP 100
           :     :     |   :| |  |   | |  :  :| |  |  |
Db      88 qqgstnrqhqrednrkskqkkinntkpea-teslivngititap 130


RESULT  27
ABG11269
ID   ABG11269 standard; Protein; 147 AA.
XX
AC   ABG11269;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #11260.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS75456.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
```

PT biodiversity -
XX
PS Claim 20; SEQ ID No 41628; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 147 AA;

```
  Query Match             13.9%;  Score 84;  DB 22;  Length 147;
  Best Local Similarity   32.9%;  Pred. No. 0.28;
  Matches   23;  Conservative    8;  Mismatches   17;  Indels   22;  Gaps    2;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQK---------------- 47
           :|:|||   ||  ||::: :|:|    |  || |    | ||:|
Db      76 kkeeeeeeeeeeeeeeeeeeee--eeeeeekkhslxdpdlgekenkgqrkrklargts 133

Qy      48 -----PGGSD 52
                |||||
Db     134 qpslspggsd 143
```

RESULT 28
ABB68123
ID ABB68123 standard; Protein; 969 AA.
XX
AC ABB68123;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 31161.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL12226.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 31161; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 969 AA;

```
  Query Match             13.9%;  Score 84;  DB 22;  Length 969;
  Best Local Similarity   27.4%;  Pred. No. 3;
  Matches   32;  Conservative   20;  Mismatches   51;  Indels   14;  Gaps    5;

Qy       1 MSQKQEEEN--PAEETGEEK-QDTQEKEGILP-----ERAEEAKLKAKYPSLGQKPGGSD 52
           :| : |||    ||  |: |   |:: :|:||     :|  |         : ||
Db      27 vssqdeeetcaeaespgqrkwtkteqsdgmlplsdkfqrirellrlefqreisqkveqla 86

Qy      53 FLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPD-----KNLVTGDHIPTPQDLP 104
             : :||.:|::  |   | :     : :|:|  | |       : : || |  || ||
Db      87 eidrrllqgrqlldrlryqvvseyyrkqqvpltgadiakvrgdslfgddiaapq-lp 142
```

RESULT 29
AAB65635
ID AAB65635 standard; Protein; 2286 AA.
XX
AC AAB65635;
XX
DT 27-MAR-2001 (first entry)
XX
DE Novel protein kinase, SEQ ID NO: 162.
XX
KW Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
KW immunosuppressive; cardiant; renal; antiinflammatory; antiasthmatic;
KW dermatological; antidiabetic; antiinfertility; gene therapy; vaccine;
KW immune disorder; cardiovascular disease; neurodegenerative disease;
KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
KW inflammatory pelvic disease; multiple sclerosis; psoriasis.
XX
OS Homo sapiens.
XX
PN WO200073469-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14842.
XX
PR 28-MAY-1999; 99US-0136503.
XX
PA (SUGE-) SUGEN INC.
XX
PI Plowman GD, Martinez R, Whyte D, Sudersanam S;
XX
DR WPI; 2001-032161/04.
DR N-PSDB; AAF44662.
XX
PT Nucleic acids encoding kinase polypeptides, useful for diagnosing and
PT treating immune-related diseases and disorders, cardiovascular disease,
PT neurodegenerative diseases and/or cancers -
XX
PS Claim 10; Fig 1; 310pp; English.

```
XX
CC   The present sequence is a novel protein kinase. The novel protein kinases
CC   and the nucleic acids that encode them may be used in the treatment and
CC   diagnosis of diseases associated with inappropriate kinase expression
CC   such as immune-related diseases and disorders, cardiovascular disease,
CC   neurodegenerative diseases and/or cancers. The nucleic acids and
CC   complementary sequences may also be used as DNA probes in diagnostic
CC   assays. The kinase polypeptides may be used as antigens in the production
CC   of antibodies of kinase expression and activity. Anti-kinase antibodies
CC   and kinase antagonists may also be used to down regulate kinase
CC   expression and activity. Diseases related to kinase expression and
CC   activity include rheumatoid arthritis, atherosclerosis, autoimmune
CC   disorders, complications of organ transplantation, myocardial infarction,
CC   immune disorders, cardiomyopathies, strokes, renal failure,
CC   oxidative-stress related disorders, chronic inflammatory bowel disease,
CC   chronic inflammatory pelvic disease, multiple sclerosis, asthma,
CC   osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
CC   reproductive disorders.
XX
SQ   Sequence   2286 AA;


  Query Match             13.9%;  Score 84;  DB 22;  Length 2286;
  Best Local Similarity   27.4%;  Pred. No. 8.8;
  Matches   29;  Conservative   16;  Mismatches   43;  Indels   18;  Gaps    4;

Qy       5 QEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKP--GGSDFLMKRLQKGQ 62
           ::|     |||       :::|  | : :||      || ||:|  | |       ::|:
Db     975 edealgtpetgaatpmdwqeqgrapsqdqeapspealpspgqepaagasp------rrge 1028

Qy      63 KYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKS 108
                 : :       || ||| : |  | |     :||||:|
Db   1029 ---------lrrgssaesalpragp-relgrglhkaasvelpqrrs 1064


RESULT  30
ABB71057
ID   ABB71057 standard; Protein; 281 AA.
XX
AC   ABB71057;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 39963.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL15160.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 39963; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   281 AA;


  Query Match             13.8%;  Score 83.5;  DB 22;  Length 281;
  Best Local Similarity   36.5%;  Pred. No. 0.72;
  Matches   27;  Conservative   12;  Mismatches   20;  Indels   15;  Gaps    4;

Qy       3 QKQEEENP-----AEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKR 57
           | ||||:|     |||  ||::| :|:|   |:       :   |:| | |||   :::
Db     120 qgqeeespdqdgeaeeaeeeeedeeeqeeatpqp------qplppalpq-pgg---vlee 169

Qy      58 LQKGQKYFDSGDYN 71
           |  ||   :  |||
Db     170 laggqvdeeledyn 183


RESULT  31
AAG16884
ID   AAG16884 standard; Protein; 374 AA.
XX
AC   AAG16884;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 17697.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
```

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.

```
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             13.8%;  Score 83.5;  DB 21;  Length 374;
  Best Local Similarity   42.2%;  Pred. No. 1;
  Matches   19;  Conservative   10;  Mismatches   15;  Indels    1;  Gaps    1;

Qy       6 EEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGG 50
           ||::  |: ||| :: :::|    |  ||:| | | ||:| | ||
Db     311 eeddidedegeedeededeeddddedeeesktkkk-psignkkgg 354


RESULT  32
AAR63437
ID   AAR63437 standard; Protein; 619 AA.
XX
AC   AAR63437;
XX
DT   19-JUL-1995  (first entry)
XX
DE   Pneumococcal surface protein A from S.pneumoniae Rx1.
XX
KW   Pneumococcal surface protein A; PspA; Streptococcus; PCR; pneumococcal;
KW   primer; protection-eliciting epitope; epitope; vaccine; amplify.
XX
OS   Streptococcus pneumoniae strain Rx1.
XX
FH   Key             Location/Qualifiers
FT   Misc_feature    192..260
FT                   /note= "protein fragment of Claim 1"
XX
PN   EP622081-A.
XX
PD   02-NOV-1994.
XX
PF   19-APR-1994;   94EP-0302767.
XX
PR   20-APR-1993;   93US-0048896.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  McDaniel LS,  Yother JL;
XX
DR   WPI; 1994-359522/45.
DR   N-PSDB; AAQ78131.
XX
PT   Epitopic regions of Pneumococcal surface protein A - derived from the
PT   Rx1 PspA strain, for the preparation of cross-reactive vaccines for
PT   the prevention of pneumococcal infections
XX
PS   Disclosure; Page 13-16; 26pp; English.
XX
CC   The aminbo acid sequence of the novel Pneumococcal surface protein A
CC   (PspA) from Streptococcus pneumoniae strain Rx1.  The gene was PCR
CC   amplified from S.pneumoniae genomic DNA using the primers AAQ78132-5.
CC   The gene was used to derive truncated peptide fragments containing
CC   protection-eliciting epitopes for use in vaccines against pneumococcal
CC   diseases.  The epitopic fragments are derived from amino acids 192-260
CC   and optionally contain a further 25 a.a. residues at both the N- and
CC   C-terminal regions of the peptide.  The epitopic peptide fragments may
CC   be derived from different strains of S.pneumoniae and are homologous to
CC   the Rx1 strain epitope.
XX
SQ   Sequence   619 AA;

  Query Match             13.7%;  Score 83;  DB 15;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 2.2;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 yntdgsmatgwlqn 427


RESULT  33
AAR87598
ID   AAR87598 standard; Protein; 619 AA.
XX
AC   AAR87598;
XX
DT   04-JUL-1996  (first entry)
XX
DE   Pneumococcal surface protein (PspA).
XX
KW   PspA; pneumococcal surface protein; truncated; immunoprotective;
KW   soluble fragment; insertion-duplication mutagenesis.
XX
OS   Streptococcus pneumoniae strain Rx1.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..31
FT                   /label= signal_peptide
FT   Protein         32..619
FT                   /label= mature_protein
FT   Region          32..319
FT                   /label= alpha-helical_coiled-coil region
FT                   /note= "contains a seven-residue periodicity"
FT   Region          320..401
FT                   /note= "proline-rich region"
FT   Region          402..421
FT                   /note= "repeat region"
FT   Region          422..441
FT                   /note= "repeat region"
FT   Region          442..461
FT                   /note= "repeat region"
FT   Region          462..481
FT                   /note= "repeat region"
FT   Region          482..501
FT                   /note= "repeat region"
FT   Region          502..521
FT                   /note= "repeat region"
FT   Region          522..541
FT                   /note= "repeat region"
FT   Region          542..561
FT                   /note= "repeat region"
FT   Region          562..581
FT                   /note= "repeat region"
FT   Region          582..619
FT                   /note= "hydrophobic region starts in last repeat
FT                         region is potential membrane-spanning region"
XX
PN   US5476929-A.
XX
PD   19-DEC-1995.
XX
PF   15-FEB-1991;   91US-0656773.
XX
PR   12-FEB-1992;   92US-0835698.
PR   15-FEB-1991;   91US-0656773.
PR   03-JUN-1993;   93US-0072070.
XX
PA   (UABR-) UAB RES FOUND.
XX
```

```
PI   Briles DE,  Mcdaniel LS,  Yother JL;
XX
DR   WPI; 1996-049021/05.
DR   N-PSDB; AAT08979.
XX
PT   New pneumococcal surface protein A fragments - comprise proline-rich
PT   region and/or repeat region, used for detection of Streptococcus
PT   pneumoniae.
XX
PS   Claim 1; Column 15-20; 23pp; English.
XX
CC   The present sequence is that of PspA (pneumococcal surface protein A)
CC   encoded by AAT08979.  Through the technique of insertion-duplication
CC   mutagenesis of the pspA gene of the strain Rx1 of Streptococcus
CC   pneumoniae with plasmids contg. cloned fragments of the pspA structural
CC   gene, it has been possible to produce soluble fragments of PspA that are
CC   secreted by pneumococci.  The method can be used to provide an
CC   immunoprotective truncated PspA protein. Primers and probes based on the
CC   present sequence are claimed, and are useful for the detection of (at
CC   least 32) S. pneumoniae strains.
XX
SQ   Sequence   619 AA;


  Query Match             13.7%;  Score 83;  DB 17;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 2.2;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 yntdgsmatgwlqn 427


RESULT  34
AAR86911
ID   AAR86911 standard; Protein; 619 AA.
XX
AC   AAR86911;
XX
DT   11-MAY-1996  (first entry)
XX
DE   Pneumococcal surface protein A.
XX
KW   Pneumococcal surface protein A; PspA; cross-protection; vaccine;
KW   Streptococcus pneumoniae; probe; primer; polymerase chain reaction;
KW   otitis media; meningitis, bacteraemia; pneumonia; epitope.
XX
OS   Streptococcus pneumoniae strain Rx1.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..31
FT                   /label= Sig_peptide
FT   Region          32..288
FT                   /note= "N-terminal region is highly charged and
FT                          includes an alpha-helix structure"
FT   Region          289..619
FT                   /note= "C-terminal region includes a proline-rich
FT                          region and a repeat region"
XX
PN   AU9520112-A.
XX
PD   30-NOV-1995.
XX
PF   18-MAY-1995;   95AU-0020112.
XX
PR   07-OCT-1994;   94US-0319795.
PR   20-MAY-1994;   94US-0246636.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  McDaniel LS,  Yother JL;
XX
DR   WPI; 1996-030801/04.
DR   N-PSDB; AAT07178.
XX
PT   Pneumococcal DNA primers and probes - amplify and detect
PT   cross-protective epitope(s) from Streptococcus pneumoniae surface
PT   protein A
XX
PS   Disclosure; Page 41-43; 61pp; English.
XX
CC   Surface protein A, PspA (AAR86911), of Streptococcus pneumoniae Rx1
CC   is the product of the pspA gene (AAT07178).  PspA includes regions,
CC   comprising e.g. amino acids 182-588, 293-588 and 192-299, that elicit
CC   cross-protection against challenge by multiple wild-type strains of
CC   S. pneumoniae.  These cross-reactive epitopes can be prepd. by
CC   expression of DNA obtd. by PCR amplification (see AAT07179-96), for
CC   use in vaccine compsns.
XX
SQ   Sequence   619 AA;


  Query Match             13.7%;  Score 83;  DB 17;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 2.2;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 yntdgsmatgwlqn 427


RESULT  35
AAY41838
ID   AAY41838 standard; Protein; 619 AA.
XX
AC   AAY41838;
XX
DT   08-DEC-1999  (first entry)
XX
DE   Streptococcus pneumoniae Rx1 PspA protein sequence.
XX
KW   Streptococcus pneumoniae Rx1; PspA; immunoprotective; vaccine;
KW   diagnosis; infection; pneumococcal surface protein A.
XX
OS   Streptococcus pneumoniae.
XX
PN   US5965400-A.
XX
PD   12-OCT-1999.
XX
PF   23-MAY-1994;   94US-0247491.
XX
PR   12-FEB-1992;   92US-0835698.
PR   15-FEB-1991;   91US-0656773.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Yother JL,  Briles DE;
XX
DR   WPI; 1999-579913/49.
DR   N-PSDB; AAZ25063.
XX
PT   DNA encoding a truncated pneumococcal surface protein A used in the
PT   development of pneumococcal infections -
XX
PS   Claim 1; Fig 3; 27pp; English.
XX
CC   The present sequence represents Streptococcus pneumoniae Rx1
```

```
CC   immunoprotective Pneumococcal surface protein A (PspA). The present
CC   invention also describes a method of forming the immunoprotective
CC   truncated PspA, comprising incorporating a vector comprising the
CC   isolated DNA molecule encoding PspA (I), into a bacterium via
CC   transformation. (I) is used to design primers which are capable of
CC   detecting a large number of S. pneumoniae strains, which in turn can
CC   be used to diagnose pneumococcal infection in mammals (e.g. humans),
CC   independent of the strain which has caused it. The PspA protein is
CC   used to develop a vaccine against pneumococcal infection comprising,
CC   as an immunologically-active component, a live attenuated or killed
CC   bacteria containing a gene coding for the truncated form of PspA.
XX
SQ   Sequence   619 AA;


  Query Match             13.7%;  Score 83;  DB 20;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 2.2;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 yntdgsmatgwlqn 427


RESULT  36
AAW70336
ID   AAW70336 standard; Protein; 648 AA.
XX
AC   AAW70336;
XX
DT   18-NOV-1998  (first entry)
XX
DE   Pneumococcal surface protein A (PspA).
XX
KW   Pneumococcal surface protein A gene; PspA; PspA epitope; vaccine;
KW   insertion-duplication mutagenesis; otitis media; meningitis;
KW   bacteraemia; pneumonia.
XX
OS   Streptococcus pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..31
FT                   /note= "Signal peptide"
FT   Protein         32..648
FT                   /note= "PspA"
FT   Region          32..319
FT                   /note= "alpha-helical coil region representing the
FT                        truncated PspA of the invention "
FT   misc_difference 647
FT                   /note= "Encoded by AGG"
XX
PN   US5804193-A.
XX
PD   08-SEP-1998.
XX
PF   17-MAR-1994;   94US-0214222.
XX
PR   12-FEB-1992;   92US-0835698.
PR   15-FEB-1991;   91US-0656773.
PR   17-MAR-1994;   94US-0214222.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  Yother JL;
XX
DR   WPI; 1998-505588/43.
DR   N-PSDB; AAV33264.
XX
PT   Truncated pneumococcal surface protein - useful in vaccines against
PT   pneumococcal infection
XX
PS   Example 3; Fig 3A-3C; 22pp; English.
XX
CC   The present sequence represents the Streptococcus pneumoniae Rx1
CC   pneumococcal surface protein A (PspA).  The invention provides a
CC   purified truncated form of PspA, formed by an insertion-duplication
CC   mutagenesis technique, comprising of the first 288 N-terminal residues of
CC   the mature form of wild-type PspA (AAW70336).  The truncated PspA
CC   contains immunoprotective epitopes of PspA.  The invention claims for a
CC   vaccine against pneumococcal infection, comprising live-attenuated or
CC   killed S. pneumoniae, containing the gene coding for the truncated PspA
CC   protein. The truncated protein, optionally conjugated to a poorly
CC   immunogenic or nonimmunogenic molecule, is claimed to be useful in
CC   vaccines against pneumococcal infection, especially otitis media,
CC   meningitis, bacteraemia and pneumonia.
XX
SQ   Sequence   648 AA;


  Query Match             13.7%;  Score 83;  DB 19;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 2.3;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 yntdgsmatgwlqn 427


RESULT  37
AAW62274
ID   AAW62274 standard; Protein; 648 AA.
XX
AC   AAW62274;
XX
DT   22-SEP-1998  (first entry)
XX
DE   Streptococcus pneumoniae pspA protein.
XX
KW   Streptococcus pneumoniae strain Rx1; pspA; immunoprotective; immunogen;
KW   pneumococcal surface protein A; cholera toxin B subunit; fusion protein;
KW   antigenic.
XX
OS   Streptococcus pneumoniae.
XX
PN   US5753463-A.
XX
PD   19-MAY-1998.
XX
PF   06-JUN-1995;   95US-0469434.
XX
PR   12-FEB-1992;   92US-0835698.
PR   15-FEB-1991;   91US-0656773.
PR   03-JUN-1993;   93US-0072065.
PR   06-JUN-1995;   95US-0469434.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  Yother JL;
XX
DR   WPI; 1998-311399/27.
DR   N-PSDB; AAV39470.
XX
PT   Truncated pneumococcal surface protein and cholera toxin B sub-unit
PT   fusion protein - useful as an immunogen against Streptococcus
PT   pneumoniae
XX
PS   Claim 1; Fig 3; 22pp; English.
XX
```

```
CC   The present sequence represents the pneumococcal surface protein A (PspA)
CC   protein from Streptococcus pneumoniae. A recombinant DNA molecule has
CC   been developed which encodes a fusion protein comprising a truncated
CC   form of PspA and cholera toxin B subunit (CTB), where the DNA molecule
CC   comprises a nucleotide sequence encoding the truncated PspA linked by
CC   an in-frame genetic fusion to a ctxB gene, and where the truncated PspA
CC   contains immunoprotective epitopes and up to 90% of the whole PspA
CC   protein, except for the cell membrane anchor region. The fusion protein
CC   is useful for providing an immunogen to protect neonates and children
CC   against S.pneumoniae. Most antigenic proteins of this strain are not
CC   immunogenic enough to provide protection. The antigenic epitopes of the
CC   fusion protein are directed against capsular polysaccharide antigens of
CC   S.pneumoniae, specifically it contains the protective epitopes of PspA.
CC   The protein can also be used in solid-phase immunoadsorbent assays,
CC   since it is readily bound to supports coated with monosialoganglioside
CC   GM1. The fusion protein is more immunogenic against S.pneumoniae than
CC   using PspA alone as the immunogen.
XX
SQ   Sequence   648 AA;


  Query Match             13.7%;  Score 83;  DB 19;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 2.3;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy      66 DSGDYNMAKAKMKN 79
           : | :||   ::|
Db     414 yntdgsmatgwlqn 427


RESULT  38
AAY41837
ID   AAY41837 standard; Protein; 648 AA.
XX
AC   AAY41837;
XX
DT   08-DEC-1999  (first entry)
XX
DE   Streptococcus pneumoniae Rx1 PspA protein sequence.
XX
KW   Streptococcus pneumoniae Rx1; PspA; immunoprotective; vaccine;
KW   diagnosis; infection; pneumococcal surface protein A.
XX
OS   Streptococcus pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 619..620
FT                   /note= "a stop codon is present in the nucleotide
FT                         sequence at this position"
FT   Misc-difference 621..622
FT                   /note= "a stop codon is present in the nucleotide
FT                         sequence at this position"
FT   Misc-difference 625..626
FT                   /note= "a stop codon is present in the nucleotide
FT                         sequence at this position"
FT   Misc-difference 630..631
FT                   /note= "a stop codon is present in the nucleotide
FT                         sequence at this position"
FT   Misc-difference 632..633
FT                   /note= "a stop codon is present in the nucleotide
FT                         sequence at this position"
XX
PN   US5965400-A.
XX
PD   12-OCT-1999.
XX
PF   23-MAY-1994;   94US-0247491.
XX
PR   12-FEB-1992;   92US-0835698.
PR   15-FEB-1991;   91US-0656773.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Yother JL,  Briles DE;
XX
DR   WPI; 1999-579913/49.
DR   N-PSDB; AAZ25063.
XX
PT   DNA encoding a truncated pneumococcal surface protein A used in the
PT   development of pneumococcal infections -
XX
PS   Claim 1; Fig 3; 27pp; English.
XX
CC   The present sequence represents Streptococcus pneumoniae Rx1
CC   immunoprotective Pneumococcal surface protein A (PspA). The present
CC   invention also describes a method of forming the immunoprotective
CC   truncated PspA, comprising incorporating a vector comprising the
CC   isolated DNA molecule encoding PspA (I), into a bacterium via
CC   transformation. (I) is used to design primers which are capable of
CC   detecting a large number of S. pneumoniae strains, which in turn can
CC   be used to diagnose pneumococcal infection in mammals (e.g. humans),
CC   independent of the strain which has caused it. The PspA protein is
CC   used to develop a vaccine against pneumococcal infection comprising,
CC   as an immunologically-active component, a live attenuated or killed
CC   bacteria containing a gene coding for the truncated form of PspA.
XX
SQ   Sequence   648 AA;


  Query Match             13.7%;  Score 83;  DB 20;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 2.3;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy      66 DSGDYNMAKAKMKN 79
           : | :||   ::|
Db     414 yntdgsmatgwlqn 427


RESULT  39
AAW87879
ID   AAW87879 standard; Protein; 648 AA.
XX
AC   AAW87879;
XX
DT   19-MAR-1999  (first entry)
XX
DE   A pneumococcal surface protein (PspA).
XX
KW   Pneumococcal surface protein; PspA; Streptococcus pneumoniae Rx1;
KW   truncated protein; epitope-containing region; vaccine;
KW   pneumococcal infection.
XX
OS   Streptococcus pneumoniae.
XX
PN   US5856170-A.
XX
PD   05-JAN-1999.
XX
PF   06-JUN-1995;   95US-0467852.
XX
PR   12-FEB-1992;   92US-0835698.
PR   15-FEB-1991;   91US-0656773.
PR   23-MAY-1994;   94US-0247491.
PR   06-JUN-1995;   95US-0467852.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  Yother JL;
```

XX
DR WPI; 1999-105118/09.
DR N-PSDB; AAV84069.
XX
PT DNA encoding truncated pneumococcal PspA protein - useful for
PT producing recombinant truncated protein
XX
PS Disclosure; Fig 3A-D; 27pp; English.
XX
CC The present sequence represents a pneumococcal surface protein (PspA)
CC of Streptococcus pneumoniae Rx1. The specification describes
CC truncated forms of PspA which contain at least the N-terminal
CC alpha-helical protective epitope-containing region. The truncated
CC PspA proteins can be used in vaccines against pneumococcal
CC infections.
XX
SQ Sequence 648 AA;

Query Match 13.7%; Score 83; DB 20; Length 648;
Best Local Similarity 31.1%; Pred. No. 2.3;
Matches 23; Conservative 13; Mismatches 36; Indels 2; Gaps 1;

Qy 8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
Db 354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy 66 DSGDYNMAKAKMKN 79
Db 414 yntdgsmatgwlqn 427

RESULT 40
AAW92456
ID AAW92456 standard; Protein; 653 AA.
XX
AC AAW92456;
XX
DT 21-APR-1999 (first entry)
XX
DE S. pneumoniae truncated pspA protein.
XX
KW PspA; pneumococcal surface protein A; immunoprotection; detection; GM1;
KW solid phase immunosorbent assay; epitope; cell membrane anchor region;
KW cholera toxin B subunit; CTB; monosialoganglioside; fusion protein.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT Misc-difference 620
FT /note= "in frame stop codon encoded by TAA"
FT Misc-difference 623
FT /note= "In frame stop codon encoded by TAA"
FT Misc-difference 628
FT /note= "In frame stop codon encoded by TAA"
FT Misc-difference 634
FT /note= "in frame stop codon encoded by TAA"
FT Misc-difference 636
FT /note= "In frame stop codon encoded by TGA"
XX
PN US5871943-A.
XX
PD 16-FEB-1999.
XX
PF 06-JUN-1995; 95US-0468718.
XX
PR 12-FEB-1992; 92US-0835698.
PR 15-FEB-1991; 91US-0656773.
PR 03-JUN-1993; 93US-0072068.
PR 06-JUN-1995; 95US-0468718.
XX
PA (UABR-) UAB RES FOUND.
XX
PI Briles DE, Yother JL;
XX
DR WPI; 1999-166635/14.
DR N-PSDB; AAX02012.
XX
PT Immunosorbent assay for pneumococcal surface protein A antigen or
PT antibody - for diagnosis of infection by Streptococcus pneumoniae
XX
PS Claim 1; Fig 3A-C; 24pp; English.
XX
CC This sequence represents a truncated form of the Streptococcus pneumoniae
CC PspA protein which is used in a solid phase immunosorbent assay for
CC detecting a PspA (pneumococcal surface protein A) antibody and antigen.
CC This truncated protein contains the immunoprotective epitopes of the
CC complete protein (up to 90% of PspA but excludes the cell-membrane
CC anchor region) fused to the B subunit of cholera toxin (CTB) which is
CC bound to monosialoganglioside (GM1) coated on the substrate. The use of
CC a fusion between truncated PspA and cholera toxin B subunit (CTB)
CC allows the support to be coated without having to isolate PspA fragments
CC since CTB binds specifically to the GM1 coating the solid support.
XX
SQ Sequence 653 AA;

Query Match 13.7%; Score 83; DB 20; Length 653;
Best Local Similarity 31.1%; Pred. No. 2.3;
Matches 23; Conservative 13; Mismatches 36; Indels 2; Gaps 1;

Qy 8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
Db 354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy 66 DSGDYNMAKAKMKN 79
Db 414 yntdgsmatgwlqn 427

Search completed: September 7, 2002, 10:45:21
Job time: 152 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:39:44 ; Search time 20.63 Seconds
(without alignments)
138.526 Million cell updates/sec

Title: US-09-824-178-3
Perfect score: 605
Sequence: 1 MSQKQEEENPAEETGEEKQD.....PTPQDLPQRKSSLVTSKLAG 117

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	451	74.5	118	2	US-08-739-484-1	Sequence 1, Appli
2	425.5	70.3	112	2	US-08-739-484-3	Sequence 3, Appli
3	414	68.4	96	2	US-08-739-484-4	Sequence 4, Appli
4	83	13.7	619	1	US-08-465-746-2	Sequence 2, Appli
5	83	13.7	619	1	US-08-214-164-2	Sequence 2, Appli
6	83	13.7	619	2	US-08-467-852A-3	Sequence 3, Appli
7	83	13.7	619	2	US-08-246-636-2	Sequence 2, Appli
8	83	13.7	619	2	US-08-247-491A-3	Sequence 3, Appli
9	83	13.7	619	2	US-08-319-795-2	Sequence 2, Appli
10	83	13.7	619	2	US-08-468-985-2	Sequence 2, Appli
11	83	13.7	619	3	US-08-312-949-2	Sequence 2, Appli
12	83	13.7	648	1	US-08-072-070-2	Sequence 2, Appli
13	83	13.7	648	1	US-08-469-434-2	Sequence 2, Appli
14	83	13.7	648	1	US-08-214-222-2	Sequence 2, Appli
15	83	13.7	648	2	US-08-467-852A-2	Sequence 2, Appli
16	83	13.7	648	2	US-08-468-718-2	Sequence 2, Appli
17	83	13.7	648	2	US-08-247-491A-2	Sequence 2, Appli
18	83	13.7	648	3	US-08-446-201-3	Sequence 3, Appli
19	83	13.7	695	1	US-08-127-499A-23	Sequence 23, Appl
20	83	13.7	695	1	US-08-482-847-23	Sequence 23, Appl
21	81.5	13.5	473	1	US-08-363-215-2	Sequence 2, Appli
22	81.5	13.5	473	3	US-08-807-342B-7	Sequence 7, Appli
23	81.5	13.5	473	4	US-08-664-962B-4	Sequence 4, Appli
24	81.5	13.5	473	4	US-09-311-743-4	Sequence 4, Appli
25	81	13.4	423	1	US-08-353-550-3	Sequence 3, Appli
26	81	13.4	423	2	US-08-551-687-3	Sequence 3, Appli
27	81	13.4	474	1	US-08-353-550-8	Sequence 8, Appli
28	81	13.4	474	2	US-08-551-687-8	Sequence 8, Appli
29	79.5	13.1	479	4	US-09-446-504-64	Sequence 64, Appl
30	79.5	13.1	479	4	US-09-712-266-64	Sequence 64, Appl
31	75.5	12.5	411	2	US-08-741-134-6	Sequence 6, Appli
32	75.5	12.5	1027	4	US-08-446-137B-2	Sequence 2, Appli
33	75	12.4	344	6	5210183-2	Patent No. 5210183
34	75	12.4	683	6	5210183-3	Patent No. 5210183
35	73.5	12.1	1250	3	US-08-938-291A-9	Sequence 9, Appli
36	73	12.1	629	3	US-08-556-419-23	Sequence 23, Appl
37	72.5	12.0	1426	4	US-09-136-574A-43	Sequence 43, Appl
38	72	11.9	564	4	US-09-308-022-6	Sequence 6, Appli
39	72	11.9	631	4	US-08-847-065-25	Sequence 25, Appl
40	72	11.9	714	2	US-08-990-114-3	Sequence 3, Appli
41	72	11.9	714	4	US-09-241-333-3	Sequence 3, Appli
42	71	11.7	312	1	US-08-285-440-4	Sequence 4, Appli
43	71	11.7	312	1	US-08-630-349-4	Sequence 4, Appli
44	71	11.7	532	1	US-08-285-440-5	Sequence 5, Appli
45	71	11.7	532	1	US-08-630-349-5	Sequence 5, Appli

ALIGNMENTS

```
RESULT   1
US-08-739-484-1
; Sequence 1, Application US/08739484
; Patent No. 5917028
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    TITLE OF INVENTION:  A NOVEL HUMAN
;    TITLE OF INVENTION:  PHOSPHOPROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/739,484
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0140 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  118 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:
;      CLONE:  Consensus
US-08-739-484-1


  Query Match             74.5%;  Score 451;  DB 2;  Length 118;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-42;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      31 RAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKN 90
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 RAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKN 87

Qy      91 LVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           |||||||||||||||||||||||||||
Db      88 LVTGDHIPTPQDLPQRKSSLVTSKLAG 114


RESULT   2
US-08-739-484-3
; Sequence 3, Application US/08739484
; Patent No. 5917028
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    TITLE OF INVENTION:  A NOVEL HUMAN
;    TITLE OF INVENTION:  PHOSPHOPROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/739,484
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0140 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  112 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  162690
US-08-739-484-3


  Query Match             70.3%;  Score 425.5;  DB 2;  Length 112;
  Best Local Similarity   76.4%;  Pred. No. 2.7e-39;
  Matches   84;  Conservative    9;  Mismatches   16;  Indels    1;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDS 67
           | |   : ||::: ::|    ||:|||||||:|| |||||||||| ||||||||||
Db       4 EVPEAASAEEQKEMEDKV-TSPEKAEEAKLKARYPHLGQKPGGSDFLRKRLQKGQKYFDS 62

Qy      68 GDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           |||||||||||||||:| |||  ||||||||||||||| ||| |||||
Db      63 GDYNMAKAKMKNKQLPTATPDKTEVTGDHIPTPQDLPQRKPSLVASKLAG 112


RESULT   3
US-08-739-484-4
; Sequence 4, Application US/08739484
; Patent No. 5917028
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    TITLE OF INVENTION:  A NOVEL HUMAN
;    TITLE OF INVENTION:  PHOSPHOPROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/739,484
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0140 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  96 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  741603
US-08-739-484-4


  Query Match             68.4%;  Score 414;  DB 2;  Length 96;
  Best Local Similarity   88.8%;  Pred. No. 4e-38;
  Matches   79;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      29 PERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPD 88
           ||:|||||||:|| |||||||||| |||||||||||||||||||||||||||||:| ||
Db       8 PEKAEEAKLKARYPHLGQKPGGSDFLRKRLQKGQKYFDSGDYNMAKAKMKNKQLPTAAPD 67

Qy      89 KNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           |  |||||||||||||||| ||| |||||
Db      68 KTEVTGDHIPTPQDLPQRKPSLVASKLAG 96


RESULT   4
US-08-465-746-2
; Sequence 2, Application US/08465746
; Patent No. 5679768
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E
;    APPLICANT:  Yother, Janet L
;    APPLICANT:  McDaniel, Larry S
;    TITLE OF INVENTION:  EPITOPIC REGIONS OF PNEUMOCOCCAL SURFACE
;    TITLE OF INVENTION:  PROTEIN A
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  Suite 1203, 2001 Jefferson Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/465,746
;       FILING DATE:
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/048,896
;       FILING DATE:
;       APPLICATION NUMBER:  US 07/656,773
;       FILING DATE:  15-FEB-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/835,698
;       FILING DATE:  12-FEB-1992
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 415-0810
;       TELEFAX:  (703) 521-0378
;       TELEX:  LUKPAT WASHINGTON
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  619 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-465-746-2


  Query Match             13.7%;  Score 83;  DB 1;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 0.5;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT   5
US-08-214-164-2
; Sequence 2, Application US/08214164
; Patent No. 5728387
;  GENERAL INFORMATION:
;    APPLICANT:  BRILES, DAVID E.
;    APPLICANT:  YOTHER, JANET L.
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  Suite 1203, 2001 Jefferson Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/214,164
;      FILING DATE:  17-MAR-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/656,773
;      FILING DATE:  15-FEB-1991
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W.
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  6102-137
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 521-0813
;      TELEX:  LUKPAT WASHINGTON
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  619 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-214-164-2


  Query Match             13.7%;  Score 83;  DB 1;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 0.5;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT   6
US-08-467-852A-3
; Sequence 3, Application US/08467852A
; Patent No. 5856170
;  GENERAL INFORMATION:
;    APPLICANT:  BRILES, David E.
;    APPLICANT:  YOTHER, Janet L.
;    APPLICANT:  MCDANIEL, Larry S.
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;      STREET:  745 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10151
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/467,852A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOWALSKI, Thomas J.
;      REGISTRATION NUMBER:  32,147
;      REFERENCE/DOCKET NUMBER:  454312-2064
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-588-0800
;      TELEFAX:  212-588-0500
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  619 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  n/a
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  amino acid
US-08-467-852A-3


  Query Match             13.7%;  Score 83;  DB 2;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 0.5;
```

```
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy      8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
          | |||:   || | |: |     |:||   :|  : |   :|| :      : : |  ||
Db    354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy     66 DSGDYNMAKAKMKN 79
          : | :||   ::|
Db    414 YNTDGSMATGWLQN 427


RESULT   7
US-08-246-636-2
; Sequence 2, Application US/08246636
; Patent No. 5965141
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E
;    APPLICANT:  Yother, Janet L
;    APPLICANT:  McDaniel, Larry S
;    APPLICANT:  Wu, Hong-Yin
;    TITLE OF INVENTION:  EPITOPIC REGIONS OF PNEUMOCOCCAL SURFACE
;    TITLE OF INVENTION:  PROTEIN A
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  Suite 1203, 2001 Jefferson Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/246,636
;      FILING DATE:  20-MAY-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/656,773
;      FILING DATE:  15-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/835,698
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/048,896
;      FILING DATE:  20-APR-1993
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;      TELEX:  LUKPAT WASHINGTON
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  619 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-246-636-2


  Query Match                 13.7%;  Score 83;  DB 2;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 0.5;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy      8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
          | |||:   || | |: |     |:||   :|  : |   :|| :      : : |  ||
Db    354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy     66 DSGDYNMAKAKMKN 79
          : | :||   ::|
Db    414 YNTDGSMATGWLQN 427


RESULT   8
US-08-247-491A-3
; Sequence 3, Application US/08247491A
; Patent No. 5965400
;  GENERAL INFORMATION:
;    APPLICANT:  BRILES, David E.
;    APPLICANT:  YOTHER, Janet L.
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;      STREET:  745 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10151
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/247,491A
;      FILING DATE:  23-JUN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOWALSKI, Thomas J.
;      REGISTRATION NUMBER:  32,147
;      REFERENCE/DOCKET NUMBER:  454312-2041
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-588-0800
;      TELEFAX:  212-588-0500
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  619 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  n/a
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  amino acid
US-08-247-491A-3


  Query Match                 13.7%;  Score 83;  DB 2;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 0.5;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy      8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
          | |||:   || | |: |     |:||   :|  : |   :|| :      : : |  ||
Db    354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy     66 DSGDYNMAKAKMKN 79
          : | :||   ::|
Db    414 YNTDGSMATGWLQN 427


RESULT   9
US-08-319-795-2
; Sequence 2, Application US/08319795
; Patent No. 5980909
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E.
;    APPLICANT:  Yother, Janet L.
;    APPLICANT:  McDaniel, Larry S
;    TITLE OF INVENTION:  Epitopic Regions of Pneumococcal Surface
;    TITLE OF INVENTION:  Protein A
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheomaker and Mattare, Ltd.
;      STREET:  1203 Crystal Plaza Bldg. 1, 2001 Jefferson
;      STREET:  Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
```

```
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/319,795
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/246,636
;      FILING DATE:  20-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/048,896
;      FILING DATE:  20-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/835,698
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/656,773
;      FILING DATE:  15-FEB-1991
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  619 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-319-795-2


  Query Match             13.7%;  Score 83;  DB 2;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 0.5;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  10
US-08-468-985-2
; Sequence 2, Application US/08468985
; Patent No. 5997882
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E.
;    APPLICANT:  Yother, Janet L.
;    APPLICANT:  McDaniel, Larry S
;    TITLE OF INVENTION:  Epitopic Regions of Pneumococcal Surface
;    TITLE OF INVENTION:  Protein A
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheomaker and Mattare, Ltd.
;      STREET:  1203 Crystal Plaza Bldg. 1, 2001 Jefferson
;      STREET:  Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/468,985
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/319,795
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/246,636
;      FILING DATE:  20-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/048,896
;      FILING DATE:  20-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/835,698
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/656,773
;      FILING DATE:  15-FEB-1991
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  619 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-468-985-2


  Query Match             13.7%;  Score 83;  DB 2;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 0.5;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  11
US-08-312-949-2
; Sequence 2, Application US/08312949
; Patent No. 6027734
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E.
;    APPLICANT:  Wu, Hong-Yin
;    TITLE OF INVENTION:  MUCOSAL ADMINISTRATION OF
;    TITLE OF INVENTION:  PNEUMOCOCCAL ANTIGENS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/312,949
;      FILING DATE:  30-SEP-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454312-2049
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  619 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-312-949-2


  Query Match             13.7%;  Score 83;  DB 3;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 0.5;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  12
US-08-072-070-2
; Sequence 2, Application US/08072070
; Patent No. 5476929
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E
;    APPLICANT:  Yother, Janet L
;    APPLICANT:  McDaniel, Larry S
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  Suite 1203, 2001 Jefferson Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/072,070
;      FILING DATE:  19930603
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/835,698
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/656,773
;      FILING DATE:  15-FEB-1991
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 521-0378
;      TELEX:  LUKPAT WASHINGTON
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  648 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-072-070-2


  Query Match             13.7%;  Score 83;  DB 1;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 0.53;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  13
US-08-469-434-2
; Sequence 2, Application US/08469434
; Patent No. 5753463
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E
;    APPLICANT:  Yother, Janet L
;    APPLICANT:  McDaniel, Larry S
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  Suite 1203, 2001 Jefferson Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,434
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/072,065
;      FILING DATE:  03 JUNE 1993
;      APPLICATION NUMBER:  US/07/835,698
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/656,773
;      FILING DATE:  15-FEB-1991
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 521-0378
;      TELEX:  LUKPAT WASHINGTON
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  648 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-469-434-2


  Query Match             13.7%;  Score 83;  DB 1;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 0.53;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  14
```

```
US-08-214-222-2
; Sequence 2, Application US/08214222
; Patent No. 5804193
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E
;    APPLICANT:  Yother, Janet L
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  Suite 1203, 2001 Jefferson Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/214,222
;      FILING DATE:  17-MAR-1994
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/835,698
;      FILING DATE:  12-FEB-1992
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 521-0378
;      TELEX:  LUKPAT WASHINGTON
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  648 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-214-222-2


  Query Match             13.7%;  Score 83;  DB 1;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 0.53;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  15
US-08-467-852A-2
; Sequence 2, Application US/08467852A
; Patent No. 5856170
;  GENERAL INFORMATION:
;    APPLICANT:  BRILES, David E.
;    APPLICANT:  YOTHER, Janet L.
;    APPLICANT:  MCDANIEL, Larry S.
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;      STREET:  745 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10151
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/467,852A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOWALSKI, Thomas J.
;      REGISTRATION NUMBER:  32,147
;      REFERENCE/DOCKET NUMBER:  454312-2064
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-588-0800
;      TELEFAX:  212-588-0500
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  648 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-467-852A-2


  Query Match             13.7%;  Score 83;  DB 2;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 0.53;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  16
US-08-468-718-2
; Sequence 2, Application US/08468718
; Patent No. 5871943
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E
;    APPLICANT:  Yother, Janet L
;    APPLICANT:  McDaniel, Larry S
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  Suite 1203, 2001 Jefferson Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/468,718
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/072,068
;      FILING DATE:  03 JUNE 1993
;      APPLICATION NUMBER:  US/07/835,698
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/656,773
;      FILING DATE:  15-FEB-1991
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
```

```
;       TELEFAX:  (703) 521-0378
;       TELEX:  LUKPAT WASHINGTON
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  648 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-468-718-2


  Query Match             13.7%;  Score 83;  DB 2;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 0.53;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
           : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  17
US-08-247-491A-2
; Sequence 2, Application US/08247491A
; Patent No. 5965400
;  GENERAL INFORMATION:
;    APPLICANT:  BRILES, David E.
;    APPLICANT:  YOTHER, Janet L.
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;      STREET:  745 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10151
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/247,491A
;      FILING DATE:  23-JUN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOWALSKI, Thomas J.
;      REGISTRATION NUMBER:  32,147
;      REFERENCE/DOCKET NUMBER:  454312-2041
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-588-0800
;      TELEFAX:  212-588-0500
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  648 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-247-491A-2


  Query Match             13.7%;  Score 83;  DB 2;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 0.53;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
           : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  18
US-08-446-201-3
; Sequence 3, Application US/08446201B
; Patent No. 6042838
; GENERAL INFORMATION:
;  APPLICANT: BRILES, David E.
;  APPLICANT: WU, Hong-Yin
;  TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS FOR MUCOSAL ADMINISTRATION OF
;  TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN A (PSPA)
;  FILE REFERENCE: 454312-2018
;  CURRENT APPLICATION NUMBER: US/08/446,201B
;  CURRENT FILING DATE: 1995-05-19
;  EARLIER APPLICATION NUMBER: 08/312,949
;  EARLIER FILING DATE: 1994-09-30
;  EARLIER APPLICATION NUMBER: 08/246,636
;  EARLIER FILING DATE: 1994-05-20
;  EARLIER APPLICATION NUMBER: 08/048,896
;  EARLIER FILING DATE: 1993-04-20
;  EARLIER APPLICATION NUMBER: 07/835,698
;  EARLIER FILING DATE: 1992-02-12
;  EARLIER APPLICATION NUMBER: 07/656,773
;  EARLIER FILING DATE: 1991-02-15
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 648
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
US-08-446-201-3


  Query Match             13.7%;  Score 83;  DB 3;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 0.53;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
           : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  19
US-08-127-499A-23
; Sequence 23, Application US/08127499A
; Patent No. 5510264
;  GENERAL INFORMATION:
;    APPLICANT:  VAN ALSTYNE, Diane
;    APPLICANT:  SHARMA, Lawrence Rajendra
;    TITLE OF INVENTION:  ANTIBODIES WHICH BIND MENINGITIS RELATED
;    TITLE OF INVENTION:  HOMOLOGOUS ANTIGENIC SEQUENCES
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/127,499A
```

```
;      FILING DATE:  28-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  51916/102/INBI
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  695 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
US-08-127-499A-23


  Query Match             13.7%;  Score 83;  DB 1;  Length 695;
  Best Local Similarity   31.1%;  Pred. No. 0.58;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     396 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 455

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     456 YNTDGSMATGWLQN 469


RESULT  20
US-08-482-847-23
; Sequence 23, Application US/08482847
; Patent No. 5556757
;  GENERAL INFORMATION:
;    APPLICANT:  VAN ALSTYNE, Diane
;    APPLICANT:  SHARMA, Lawrence Rajendra
;    TITLE OF INVENTION:  PEPTIDES REPRESENTING EPITOPIC SITES FOR
;    TITLE OF INVENTION:  BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR
;    TITLE OF INVENTION:  CNS CARRIER, ANTIBODIES THERETO, AND USES THEREOF
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/482,847
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/127,499
;      FILING DATE:  28-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  51916/104/INBI
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  695 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
US-08-482-847-23


  Query Match             13.7%;  Score 83;  DB 1;  Length 695;
  Best Local Similarity   31.1%;  Pred. No. 0.58;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     396 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 455

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     456 YNTDGSMATGWLQN 469


RESULT  21
US-08-363-215-2
; Sequence 2, Application US/08363215
; Patent No. 5807989
;  GENERAL INFORMATION:
;    APPLICANT:  Margolis, Benjamin L.
;    APPLICANT:  Schlessinger, Joseph
;    APPLICANT:  Yajnik, Vijay
;    TITLE OF INVENTION:  METHODS FOR TREATMENT OR
;    TITLE OF INVENTION:  DIAGNOSIS OF DISEASES OR
;    TITLE OF INVENTION:  DISORDERS ASSOCIATED
;    TITLE OF INVENTION:  WITH AN APB DOMAIN
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/363,215
;      FILING DATE:  December 23, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  212/208
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  473 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-363-215-2


  Query Match             13.5%;  Score 81.5;  DB 1;  Length 473;
  Best Local Similarity   26.2%;  Pred. No. 0.52;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   33;  Gaps    7;
```

```
Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :         |:|        | : :|| |  |       |: ||        :||
Db     229 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 288

Qy      47 KPGGSDFLMKRLQ-----KGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDHI 97
             ||    : |::        |:: ||    |     |: ||:    : :  |||     : |
Db     289 PVGGDPEVRKQMPPPPPCPGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS-- 343

Qy      98 PTPQDL 103
             |:||
Db     344 -APRDL 348


RESULT  22
US-08-807-342B-7
; Sequence 7, Application US/08807342B
; Patent No. 6077686
;  GENERAL INFORMATION:
;    APPLICANT:  Der, Channing
;    APPLICANT:  O'Bryan, John P.
;    APPLICANT:  Pawson, Anthony
;    TITLE OF INVENTION:  No. 6077686el SHC Proteins
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr., P.O. Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/807,342B
;      FILING DATE:  28-FEB-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/012,516
;      FILING DATE:  29-FEB-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  MTS2USA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9200
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  473 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-807-342B-7


  Query Match             13.5%;  Score 81.5;  DB 3;  Length 473;
  Best Local Similarity   26.2%;  Pred. No. 0.52;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   33;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :         |:|        | : :|| |  |       |: ||        :||
Db     229 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 288

Qy      47 KPGGSDFLMKRLQ-----KGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDHI 97
             ||    : |::        |:: ||    |     |: ||:    : :  |||     : |
Db     289 PVGGDPEVRKQMPPPPPCPGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS-- 343

Qy      98 PTPQDL 103
             |:||
Db     344 -APRDL 348


RESULT  23
US-08-664-962B-4
; Sequence 4, Application US/08664962B
; Patent No. 6218162
;  GENERAL INFORMATION:
;    APPLICANT:  Krystal, Gerald
;    TITLE OF INVENTION:  SH2-CONTAINING INOSITOL-PHOSPHATASE
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MERCHANT & GOULD
;      STREET:  3100 No. 6218162west Center, 90 South Seventh Street
;      CITY:  Minneapolis
;      STATE:  Minnesota
;      COUNTRY:  U.S.A.
;      ZIP:  55402-4131
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/664,962B
;      FILING DATE:  14-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mueller, Douglas P.
;      REGISTRATION NUMBER:  30,300
;      REFERENCE/DOCKET NUMBER:  M&G 7933.49-US-01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612-332-5300
;      TELEFAX:  612-332-9081
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 473 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-664-962B-4


  Query Match             13.5%;  Score 81.5;  DB 4;  Length 473;
  Best Local Similarity   26.2%;  Pred. No. 0.52;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   33;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :         |:|        | : :|| |  |       |: ||        :||
Db     229 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 288

Qy      47 KPGGSDFLMKRLQ-----KGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDHI 97
             ||    : |::        |:: ||    |     |: ||:    : :  |||     : |
Db     289 PVGGDPEVRKQMPPPPPCPGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS-- 343

Qy      98 PTPQDL 103
             |:||
Db     344 -APRDL 348


RESULT  24
US-09-311-743-4
; Sequence 4, Application US/09311743
; Patent No. 6238903
;  GENERAL INFORMATION:
;       APPLICANT: Krystal, Gerald
;       TITLE OF INVENTION: SH2-CONTAINING INOSITOL-PHOSPHATASE
;       NUMBER OF SEQUENCES: 24
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: BERESKIN & PARR
;            STREET: 40 King Street West
;            CITY: Toronto
```

```
;              STATE: Ontario
;              COUNTRY: Canada
;              ZIP: M5H 3Y2
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/311,743
;              FILING DATE: 14-May-1999
;              CLASSIFICATION: <Unknown>
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Gravelle, Micheline
;              REGISTRATION NUMBER: 40,261
;              REFERENCE/DOCKET NUMBER: 7771-32
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 416-364-7311
;              TELEFAX: 416-361-1398
;    INFORMATION FOR SEQ ID NO: 4:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 473 amino acids
;              TYPE: amino acid
;              TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-311-743-4


  Query Match             13.5%;  Score 81.5;  DB 4;  Length 473;
  Best Local Similarity   26.2%;  Pred. No. 0.52;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   33;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|       | : :||  |  |       |: ||       :||
Db     229 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 288

Qy      47 KPGGSDFLMKRLQ-----KGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDHI 97
             ||   : |::       |:: ||   |    |: ||:   : :  |||    : |
Db     289 PVGGDPEVRKQMPPPPPCPGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS-- 343

Qy      98 PTPQDL 103
             |:||
Db     344 -APRDL 348


RESULT  25
US-08-353-550-3
; Sequence 3, Application US/08353550
; Patent No. 5744313
;  GENERAL INFORMATION:
;    APPLICANT:  Williams, Lewis T.
;    APPLICANT:  Cavanaugh, William M.
;    TITLE OF INVENTION:  No. 5744313el Protein Domain Which Binds
;    TITLE OF INVENTION:  Tyrosine Phosphorylated Proteins
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/353,550
;      FILING DATE:  09-DEC-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dow, Karen B.
;      REGISTRATION NUMBER:  29,684
;      REFERENCE/DOCKET NUMBER:  02307K-057300
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-353-550-3


  Query Match             13.4%;  Score 81;  DB 1;  Length 423;
  Best Local Similarity   26.0%;  Pred. No. 0.51;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   34;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|       | : :||  |  |       |: ||       :||
Db     219 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 278

Qy      47 KPGGSDFLMKRL------QKGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDH 96
             ||   : |::         |:: ||   |    |: ||:   : :  |||    : |
Db     279 PVGGDPEVRKQMPPPPPCPAGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS- 334

Qy      97 IPTPQDL 103
             |:||
Db     335 --APRDL 339


RESULT  26
US-08-551-687-3
; Sequence 3, Application US/08551687
; Patent No. 5925547
;  GENERAL INFORMATION:
;    APPLICANT:  Williams, Lewis T.
;    APPLICANT:  Cavanaugh, William M.
;    TITLE OF INVENTION:  No. 5925547el Protein Domain Which Binds
;    TITLE OF INVENTION:  Tyrosine Phosphorylated Proteins
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/551,687
;      FILING DATE:  01-NOV-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/353,550
;      FILING DATE:  09-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy, Matthew B.
;      REGISTRATION NUMBER:  P39,787
;      REFERENCE/DOCKET NUMBER:  2307K-5731
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
```

```
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-551-687-3


  Query Match             13.4%;  Score 81;  DB 2;  Length 423;
  Best Local Similarity   26.0%;  Pred. No. 0.51;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   34;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|       | : :||  |  |       |: ||       :||
Db     219 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 278

Qy      47 KPGGSDFLMKRL------QKGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDH 96
             ||   : |::        |:: ||   |    |: ||:   : :  |||    : |
Db     279 PVGGDPEVRKQMPPPPPCPAGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS- 334

Qy      97 IPTPQDL 103
             |:||
Db     335 --APRDL 339


RESULT  27
US-08-353-550-8
; Sequence 8, Application US/08353550
; Patent No. 5744313
;  GENERAL INFORMATION:
;    APPLICANT:  Williams, Lewis T.
;    APPLICANT:  Cavanaugh, William M.
;    TITLE OF INVENTION:  No. 5744313el Protein Domain Which Binds
;    TITLE OF INVENTION:  Tyrosine Phosphorylated Proteins
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/353,550
;      FILING DATE:  09-DEC-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dow, Karen B.
;      REGISTRATION NUMBER:  29,684
;      REFERENCE/DOCKET NUMBER:  02307K-057300
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  474 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-353-550-8


  Query Match             13.4%;  Score 81;  DB 1;  Length 474;
  Best Local Similarity   26.0%;  Pred. No. 0.59;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   34;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|       | : :||  |  |       |: ||       :||
Db     229 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 288

Qy      47 KPGGSDFLMKRL------QKGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDH 96
             ||   : |::        |:: ||   |    |: ||:   : :  |||    : |
Db     289 PVGGDPEVRKQMPPPPPCPAGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS- 344

Qy      97 IPTPQDL 103
             |:||
Db     345 --APRDL 349


RESULT  28
US-08-551-687-8
; Sequence 8, Application US/08551687
; Patent No. 5925547
;  GENERAL INFORMATION:
;    APPLICANT:  Williams, Lewis T.
;    APPLICANT:  Cavanaugh, William M.
;    TITLE OF INVENTION:  No. 5925547el Protein Domain Which Binds
;    TITLE OF INVENTION:  Tyrosine Phosphorylated Proteins
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/551,687
;      FILING DATE:  01-NOV-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/353,550
;      FILING DATE:  09-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy, Matthew B.
;      REGISTRATION NUMBER:  P39,787
;      REFERENCE/DOCKET NUMBER:  2307K-5731
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  474 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-551-687-8


  Query Match             13.4%;  Score 81;  DB 2;  Length 474;
  Best Local Similarity   26.0%;  Pred. No. 0.59;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   34;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|       | : :||  |  |       |: ||       :||
Db     229 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 288

Qy      47 KPGGSDFLMKRL------QKGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDH 96
             ||   : |::        |:: ||   |    |: ||:   : :  |||    : |
Db     289 PVGGDPEVRKQMPPPPPCPAGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS- 344

Qy      97 IPTPQDL 103
             |:||
Db     345 --APRDL 349


RESULT  29
US-09-446-504-64
```

```
; Sequence 64, Application US/09446504
; Patent No. 6218150
; GENERAL INFORMATION:
;  APPLICANT:  UEMORI, Takashi
;  APPLICANT:  SATO, Yoshimi
;  APPLICANT:  FUJITA, Tomoko
;  APPLICANT:  MIYAKE, Kazue
;  APPLICANT:  MUKAI, Hiroyuki
;  APPLICANT:  ASADA, Kiyozo
;  APPLICANT:  KATO, Ikunoshin
;  TITLE OF INVENTION:  DNA POLYMERASE-RELATED FACTORS
;  FILE REFERENCE:  1422-408PCT
;  CURRENT APPLICATION NUMBER: US/09/446,504
;  CURRENT FILING DATE:  1999-12-23
;  PRIOR APPLICATION NUMBER:  PCT/JP98/02845
;  PRIOR FILING DATE:  1998-06-24
;  PRIOR APPLICATION NUMBER:  JP 9-187496
;  PRIOR FILING DATE:  1997-06-26
;  PRIOR APPLICATION NUMBER:  JP 9-320692
;  PRIOR FILING DATE:  1997-11-27
;  NUMBER OF SEQ ID NOS:  92
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 64
;   LENGTH: 479
;   TYPE: PRT
;   ORGANISM: Pyrococcus furiosus
US-09-446-504-64


  Query Match             13.1%;  Score 79.5;  DB 4;  Length 479;
  Best Local Similarity   38.5%;  Pred. No. 0.87;
  Matches   20;  Conservative    9;  Mismatches   20;  Indels    3;  Gaps    1;

Qy       5 QEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMK 56
           :|||   |:| |||:: :||     |  || | : : |  |::    ||| |
Db     431 EEEEEEEEKTEEEKEEIEEKP---EEEKEEEKKEKEKPKKGKQATLFDFLKK 479


RESULT  30
US-09-712-266-64
; Sequence 64, Application US/09712266
; Patent No. 6333158
; GENERAL INFORMATION:
;  APPLICANT: UEMORI, Takashi
;  APPLICANT:  SATO, Yoshimi
;  APPLICANT:  FUJITA, Tomoko
;  APPLICANT:  MIYAKE, Kazue
;  APPLICANT:  MUKAI, Hiroyuki
;  APPLICANT:  ASADA, Kiyozo
;  APPLICANT:  KATO, Ikunoshin
;  TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
;  FILE REFERENCE: 1422-408PCT
;  CURRENT APPLICATION NUMBER: US/09/712,266
;  CURRENT FILING DATE:  2000-11-15
;  PRIOR APPLICATION NUMBER: US 09/446,504
;  PRIOR FILING DATE: 1999-12-23
;  PRIOR APPLICATION NUMBER: PCT/JP98/02845
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: JP 9-187496
;  PRIOR FILING DATE: 1997-06-26
;  PRIOR APPLICATION NUMBER: JP 9-320692
;  PRIOR FILING DATE: 1997-11-27
;  NUMBER OF SEQ ID NOS: 92
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
;   LENGTH: 479
;   TYPE: PRT
;   ORGANISM: Pyrococcus furiosus
US-09-712-266-64


  Query Match             13.1%;  Score 79.5;  DB 4;  Length 479;
  Best Local Similarity   38.5%;  Pred. No. 0.87;
  Matches   20;  Conservative    9;  Mismatches   20;  Indels    3;  Gaps    1;

Qy       5 QEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMK 56
           :|||   |:| |||:: :||     |  || | : : |  |::    ||| |
Db     431 EEEEEEEEKTEEEKEEIEEKP---EEEKEEEKKEKEKPKKGKQATLFDFLKK 479


RESULT  31
US-08-741-134-6
; Sequence 6, Application US/08741134
; Patent No. 5861498
;  GENERAL INFORMATION:
;    APPLICANT:  Litwack, Gerald
;    APPLICANT:  Alnemri, Emad S.
;    APPLICANT:  Fernandes-Alnemri, Teresa
;    TITLE OF INVENTION:  IMMUNOPHILIN FKBP46 AND COMPOSITIONS FOR MAKING
;    TITLE OF INVENTION:  AND
;    TITLE OF INVENTION:  METHODS OF USING THE SAME
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5861498ris
;      STREET:  One Liberty Place - 46th floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Windows 3.11
;      SOFTWARE:  WordPerfect for Windows 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/741,134
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/007,163
;      FILING DATE:  01-NOV-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca, Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  TJU-2090
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 411 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-741-134-6


  Query Match             12.5%;  Score 75.5;  DB 2;  Length 411;
  Best Local Similarity   25.2%;  Pred. No. 2;
  Matches   28;  Conservative   20;  Mismatches   52;  Indels   11;  Gaps    2;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           : :: ||:|  ||  ||:::: ::|| : ||  :  | | :     ::        |: :|
Db     219 VQEEDEEDNDGEEEQEEEEEEEQKEEVKPEPKKSKKEKKRKHEEKEEE-------KKAKK 271

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHI----PTPQDLPQRK 107
           :|     |      | |:|:       |: |    |     |  | || |
Db     272 VKKVEFKKDLEEGPTKPKSKKEQDKHKPKSKVLEGGIVIEDRTIGDGPQAK 322


RESULT  32
US-08-446-137B-2
; Sequence 2, Application US/08446137B
; Patent No. 6162903
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Trowern, Angus R.
;    APPLICANT:  Atkinson, Anthony
;    APPLICANT:  Murphy, Jonathan P.
;    APPLICANT:  Laurence, Oliver S.
;    APPLICANT:  Duggleby, Clive J.
;    TITLE OF INVENTION:  IMMUNOGLOBULIN BINDING PROTEINS DERIVED
;    TITLE OF INVENTION:  FROM L PROTEIN AND THEIR USES
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,137B
;      FILING DATE:  22-MAY-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
;      REFERENCE/DOCKET NUMBER:  100084.406
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1027 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-446-137B-2


  Query Match             12.5%;  Score 75.5;  DB 4;  Length 1027;
  Best Local Similarity   25.8%;  Pred. No. 6.3;
  Matches   23;  Conservative   19;  Mismatches   28;  Indels   19;  Gaps    3;

Qy       5 QEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKY 64
           :::| | :: ||:|:   :| |      |: | : | |   :||       :  : |:
Db     938 KKDEEPGKKPGEDKKPEDKKPG------EDKKPEDKKPGEDKKP-------EDKKPGKTD 984

Qy      65 FDSGDYNMAKAKMKNKQLPSAGPDKNLVT 93
             || :      | |  :|| || :  ::|
Db     985 KDSPN------KKKKAKLPKAGSEAEILT 1007


RESULT  33
5210183-2
;Patent No. 5210183
;    APPLICANT: LINDAHL, GUNNAR;FRITHZ, ELISABET;HEDEN, LARS-OLOF
;    TITLE OF INVENTION: PROTEIN ARP, WITH IMMUNOGLOBULIN A
;BINDING ACTIVITY, THE CORRESPONDING VECTORS AND HOSTS, REAGENT
;KIT AND PHARMACEUTICAL COMPOSITION
;    NUMBER OF SEQUENCES: 3
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/270,099
;      FILING DATE: 14-NOV-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 186,097
;      FILING DATE: 25-APR-1988
;SEQ ID NO:2:
;      LENGTH: 344
5210183-2


 Query Match             12.4%;  Score 75;  DB 6;  Length 344;
  Best Local Similarity   24.4%;  Pred. No. 1.8;
  Matches   30;  Conservative   24;  Mismatches   57;  Indels   12;  Gaps    4;

Qy       4 KQEEENPAEETGEEKQ-DTQEKEGILPE----RAEEAKLKAKYPSLGQKPGGSDFLMKRL 58
           |:| |  |::  ||||     ::|:  :    |  : |::|        |    : | | |
Db     185 KKELEAEAQKLKEEKQISDASRQGLSRDLEASRERKKKVEADLAEANSKLQALEKLNKEL 244

Qy      59 QKGQKYFDSGDYNM-----AKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQ--RKSSLV 111
           ::|:|  :     :      |:||   :||      :  : |:  |  :  ||  |  | :
Db     245 EEGKKLSEKEKAELQARLEAEAKALKEQLAKQAEELAKLKGNQTPNAKVAPQANRSRSAM 304

Qy     112 TSK 114
           | :
Db     305 TQQ 307


RESULT  34
5210183-3
;Patent No. 5210183
;    APPLICANT: LINDAHL, GUNNAR;FRITHZ, ELISABET;HEDEN, LARS-OLOF
;    TITLE OF INVENTION: PROTEIN ARP, WITH IMMUNOGLOBULIN A
;BINDING ACTIVITY, THE CORRESPONDING VECTORS AND HOSTS, REAGENT
;KIT AND PHARMACEUTICAL COMPOSITION
;    NUMBER OF SEQUENCES: 3
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/270,099
;      FILING DATE: 14-NOV-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 186,097
;      FILING DATE: 25-APR-1988
;SEQ ID NO:3:
;      LENGTH: 683
5210183-3


  Query Match             12.4%;  Score 75;  DB 6;  Length 683;
  Best Local Similarity   24.4%;  Pred. No. 4.2;
  Matches   30;  Conservative   24;  Mismatches   57;  Indels   12;  Gaps    4;

Qy       4 KQEEENPAEETGEEKQ-DTQEKEGILPE----RAEEAKLKAKYPSLGQKPGGSDFLMKRL 58
           |:| |  |::  ||||     ::|:  :    |  : |::|        |    : | | |
Db     185 KKELEAEAQKLKEEKQISDASRQGLSRDLEASRERKKKVEADLAEANSKLQALEKLNKEL 244

Qy      59 QKGQKYFDSGDYNM-----AKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQ--RKSSLV 111
           ::|:|  :     :      |:||   :||      :  : |:  |  :  ||  |  | :
Db     245 EEGKKLSEKEKAELQARLEAEAKALKEQLAKQAEELAKLKGNQTPNAKVAPQANRSRSAM 304

Qy     112 TSK 114
           | :
Db     305 TQQ 307


RESULT  35
US-08-938-291A-9
; Sequence 9, Application US/08938291A
; Patent No. 6117673
;  GENERAL INFORMATION:
;    APPLICANT:  Lev, Sima
;    APPLICANT:  Plowman, Gregory D.
;    APPLICANT:  Schlessinger, Joseph
;    TITLE OF INVENTION:  RDGB PROTEINS AND RELATED
;    TITLE OF INVENTION:  PRODUCTS AND METHODS
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
```

```
;       MEDIUM TYPE:  storage
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  IBM P.C. DOS 5.0
;       SOFTWARE:  FastSeq
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/938,291A
;       FILING DATE:  September 26, 1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  60/027,337
;       FILING DATE:  October 11, 1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Warburg, Richard J.
;       REGISTRATION NUMBER:  32,327
;       REFERENCE/DOCKET NUMBER:  228/172
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1250 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-08-938-291A-9


  Query Match             12.1%;  Score 73.5;  DB 3;  Length 1250;
  Best Local Similarity   23.8%;  Pred. No. 13;
  Matches   34;  Conservative   18;  Mismatches   50;  Indels   41;  Gaps    6;

Qy       3 QKQEEENPAEETGEEKQDTQEKE-------------GILPERAEEAKL----KAKYPSL 44
           |    |  ::| |||::|  | |              |   :|:    :    | |: |
Db     311 QPPSAEASSDEEGEEEEDDDEDENDAIGTGVDLSANQGGSAQRSRSQSIQMAQKGKFGSK 370

Qy      45 G--QKPGGS---------DFLMKRLQ------KGQKYFDSGDYNMAKAKMKNKQLPSAGP 87
           |    | ||         :: |:||:        :::||  | |   :  |   |   |
Db     371 GALHSPVGSAHSFDLQVANWRMERLEVDSKSNSDEEFFDCLDTNETNSLAKWSSLELLG- 429

Qy      88 DKNLVTGDHIPTPQDLPQRKSSL 110
             ||  | |   |   :|:
Db     430 -----EGDDSPPPHGGPSSAASV 447


RESULT  36
US-08-556-419-23
; Sequence 23, Application US/08556419C
; Patent No. 6093549
; GENERAL INFORMATION:
;  APPLICANT: Ross, Christopher
;  APPLICANT: Li, Xiao-Jiang
;  APPLICANT: Li, Shi-Hua
;  APPLICANT: Sharp, Alan
;  APPLICANT: Lanahan, Anthony
;  APPLICANT: Worley, Paul
;  APPLICANT: Snyder, Solomon
;  TITLE OF INVENTION: Huntingtin-associated protein
;  FILE REFERENCE: 01107.52271
;  CURRENT APPLICATION NUMBER: US/08/556,419C
;  CURRENT FILING DATE: 1995-11-09
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 23
;   LENGTH: 629
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-08-556-419-23


  Query Match             12.1%;  Score 73;  DB 3;  Length 629;
  Best Local Similarity   25.8%;  Pred. No. 6.3;
  Matches   24;  Conservative   18;  Mismatches   43;  Indels    8;  Gaps    3;

Qy       7 EENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL---GQKPGGSD--FLMKRLQKG 61
           |: ||||| :  ::|:  | : ||  |  ::     :|   |  |    |  :::::|
Db     503 EDGPAEETEQASEETEAWEEVEPEVDEATRMNVVVSALEASGLGPSHLDMKYVLQQLSNW 562

Qy      62 QKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTG 94
           |   |:      | |:  |  |:     |:  |
Db     563 Q---DAHSKRQQKQKVVPKDSPAPQQQTNMGGG 592


RESULT  37
US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366
;   GENERAL INFORMATION:
;       APPLICANT: Farrington, Graham K.
;                  Anderson, Paige
;                  Gibbs, Moreland
;                  Bergquist, Peter
;                  Daniels, Roy
;                  Morgan, Hugh W.
;                  Williams, Diane P.
;       TITLE OF INVENTION: Compositions and Methods for
;                           Treating Cellulose Containing Fabrics Using Truncated
;                           Cellulase Enzyme Compositions
;       NUMBER OF SEQUENCES: 49
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Howson and Howson
;            STREET: Spring House Corporate Center, P.O. Box 457
;            CITY: Spring House
;            STATE: PA
;            COUNTRY: USA
;            ZIP: 19477
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: DOS
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/136,574A
;            FILING DATE: 19-Aug-1998
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US 08/932,571
;            FILING DATE: September 19, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Bak, Mary E.
;            REGISTRATION NUMBER: 31,215
;            REFERENCE/DOCKET NUMBER: 1997US001/CIP
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 215-540-9200
;            TELEFAX: 215-540-5818
;            TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 43:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1426 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: No. 6294366e
;       SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-09-136-574A-43


  Query Match             12.0%;  Score 72.5;  DB 4;  Length 1426;
  Best Local Similarity   23.4%;  Pred. No. 20;
  Matches   30;  Conservative   16;  Mismatches   39;  Indels   43;  Gaps    6;

Qy      15 GEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMK----------------- 56
           ||  |       | |     | :  |:  ||   ||:|: ::
Db     827 GEATQSVSVTSSINP-----AYIDVKFVKLGANAGGADYYVEIGFKSGAGVLAAGQSTKE 881

Qy      57 ---RLQKGQ-KYFDSGDYNM--AKAKMKNKQL------------PSAG-PDKNLVTGDH 96
```

```
              :|||   |  | ||::  | : ::|:::            || |     :||
Db    882 IRLSIQKGSGSYNQSNDYSIRSANSYIENEKVTGYIDGAIVWGREPSRGTKPAGVVTPTP 941

Qy     97 IPTPQDLP 104
          |||   |
Db    942 APTPTSTP 949
```

RESULT 38
US-09-308-022-6
; Sequence 6, Application US/09308022
; Patent No. 6291654
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA, ET AL.
; TITLE OF INVENTION: C3 BINDING PROTEIN OF STREPTOCOCCUS
; TITLE OF INVENTION: PNEUMONIAE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MUETING, RAASCH & GEBHARDT, P.A.
; STREET: 119 No. 6291654th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: Minnesota
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/308,022
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/20586
; FILING DATE: 12-NOV-1997
; APPLICATION NUMBER: 60/029,444
; FILING DATE: 12-NOV-1996
; APPLICATION NUMBER: 60/038,086
; FILING DATE: 18-FEB-1997
; APPLICATION NUMBER: 60/059,368
; FILING DATE: 19-SEP-1997
; APPLICATION NUMBER: 60/062,473
; FILING DATE: 16-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: MUETING, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 110.00430101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 564 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-308-022-6

Query Match 11.9%; Score 72; DB 4; Length 564;
Best Local Similarity 30.3%; Pred. No. 7.1;
Matches 23; Conservative 13; Mismatches 36; Indels 4; Gaps 2;

```
Qy      8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPG--GSDFLMKRLQKGQK 63
          | |||:   || | |: |     |:||   :|  : |   :||    :     : : |
Db    433 EKPAEQPKAEKTDDQQAEEDYARRSEEEYNRLTQQQPPKTEKPAQPSTPKTGWKQENGMW 492

Qy     64 YFDSGDYNMAKAKMKN 79
          || : | :||   ::|
Db    493 YFYNTDGSMATGWLQN 508
```

RESULT 39
US-08-847-065-25
; Sequence 25, Application US/08847065
; Patent No. 6245335
; GENERAL INFORMATION:
; APPLICANT: Masure, H. Robert
; APPLICANT: Rosenow, Carsten I.
; APPLICANT: Tuomanen, Elaine
; APPLICANT: Wizemann, Theresa M.
; TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR
; TITLE OF INVENTION: ANTI-PNEUMOCOCCAL VACCINES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/847,065
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-158 ..
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 631 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
US-08-847-065-25

Query Match 11.9%; Score 72; DB 4; Length 631;
Best Local Similarity 30.3%; Pred. No. 8.1;
Matches 23; Conservative 13; Mismatches 36; Indels 4; Gaps 2;

```
Qy      8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPG--GSDFLMKRLQKGQK 63
          | |||:   || | |: |     |:||   :|  : |   :||    :     : : |
Db    362 EKPAEQPKAEKTDDQQAEEDYARRSEEEYNRLTQQQPPKTEKPAQPSTPKTGWKQENGMW 421

Qy     64 YFDSGDYNMAKAKMKN 79
          || : | :||   ::|
Db    422 YFYNTDGSMATGWLQN 437
```

RESULT 40
US-08-990-114-3
; Sequence 3, Application US/08990114
; Patent No. 5932475
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN NUCLEOLIN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 3

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/990,114
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0451 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  714 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  128842
US-08-990-114-3


  Query Match             11.9%;  Score 72;  DB 2;  Length 714;
  Best Local Similarity   26.3%;  Pred. No. 9.5;
  Matches   26;  Conservative   13;  Mismatches   26;  Indels   34;  Gaps    4;

Qy       5 QEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKY 64
           :||:   ||  ||::| :|:|    |  ||  :|   |: |::
Db     246 EEEDEDEEEDEEEEEDEEEE----EEEEEEPVK---PAPGKR---------------- 281

Qy      65 FDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDL 103
                |   | |::| |   |  | |    || :|
Db     282 --------KKEMTKQKEVPEA--KKQKVEGSESTTPFNL 310


Search completed: September  7, 2002, 10:43:49
Job time: 245 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:41:44 ; Search time 28.24 Seconds
(without alignments)
398.104 Million cell updates/sec

Title: US-09-824-178-3
Perfect score: 605
Sequence: 1 MSQKQEEENPAEETGEEKQD.....PTPQDLPQRKSSLVTSKLAG 117

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	425.5	70.3	112	2	B35308	cAMP-regulated pho
2	406	67.1	77	2	JC4878	alpha endosulfine
3	128.5	21.2	174	2	T23557	hypothetical prote
4	122	20.2	139	2	T37510	hypothetical prote
5	95.5	15.8	842	2	G90576	heat shock atp-dep
6	91.5	15.1	2664	2	T28626	variant-specific s
7	90.5	15.0	168	2	S60970	hypothetical prote
8	85	14.0	775	2	A84526	hypothetical prote
9	83.5	13.8	407	1	EDBEQ3	immediate-early pr
10	83	13.7	619	2	A41971	surface protein ps
11	83	13.7	619	2	A97887	surface protein ps
12	83	13.7	763	2	T08929	hypothetical prote
13	83	13.7	792	2	E84525	hypothetical prote
14	82.5	13.6	795	2	D84477	hypothetical prote
15	82	13.6	644	2	S55395	neurofilament prot
16	81.5	13.5	473	2	S25776	transforming prote
17	81.5	13.5	646	2	T02398	hypothetical prote
18	81.5	13.5	789	2	E84514	hypothetical prote
19	81	13.4	155	2	JQ0797	histone H2B.IV - V
20	81	13.4	325	2	I38616	zinc finger protei
21	81	13.4	695	2	I51652	dsRNA-binding prot
22	80.5	13.3	522	2	C96608	hypothetical prote
23	80.5	13.3	820	2	T12972	hypothetical prote
24	80	13.2	215	2	T23482	hypothetical prote
25	80	13.2	751	2	S65469	DNA topoisomerase
26	79	13.1	335	2	T33457	hypothetical prote
27	79	13.1	785	2	C88131	protein F10G7.1 [i
28	79	13.1	858	2	S15762	neurofilament trip
29	79	13.1	1014	2	T31433	Na+/Ca2+,K+-exchan
30	79	13.1	4967	2	S72269	ryanodine receptor
31	78.5	13.0	260	2	S60480	low temperature-in
32	78.5	13.0	1359	2	T34036	hypothetical prote
33	78	12.9	535	2	T41384	hypothetical prote
34	78	12.9	742	2	T48930	hypothetical prote
35	78	12.9	792	2	H83736	subtilisin-type al
36	77.5	12.8	1090	2	H86806	hypothetical prote
37	77.5	12.8	1351	2	C71607	hypothetical prote
38	77.5	12.8	2109	2	E89066	protein H05009.1 [
39	77.5	12.8	2109	2	T33247	hypothetical prote
40	77	12.7	366	2	AC1261	hypothetical prote
41	77	12.7	366	2	AE1623	hypothetical prote
42	77	12.7	784	2	T45027	hypothetical prote
43	77	12.7	1132	2	T43483	translation initia
44	77	12.7	1738	2	T14867	interaptin - slime
45	76.5	12.6	274	2	S16925	caldesmon - turkey

ALIGNMENTS

RESULT 1
B35308
cAMP-regulated phosphoprotein, 19K - bovine
N;Alternate names: ARPP-19
N;Contains: cAMP-regulated phosphoprotein, 16K
C;Species: Bos primigenius taurus (cattle)
C;Date: 17-Aug-1990 #sequence_revision 17-Aug-1990 #text_change 24-Nov-1999
C;Accession: B35308; A35308
R;Horiuchi, A.; Williams, K.R.; Kurihara, T.; Nairn, A.C.; Greengard, P.
J. Biol. Chem. 265, 9476-9484, 1990
A;Title: Purification and cDNA cloning of ARPP-16, a cAMP-regulated phosphoprotein en
A;Reference number: A35308; MUID:90264450
A;Accession: B35308
A;Molecule type: mRNA
A;Residues: 1-112 <HOR>
A;Cross-references: GB:M33618; NID:g162689; PIDN:AAA30386.1; PID:g162690; GB:J05450
A;Accession: A35308
A;Molecule type: mRNA
A;Residues: 17-112 <HO2>
A;Cross-references: GB:M33617; NID:g162687; PIDN:AAA30385.1; PID:g162688; GB:J05450
C;Keywords: acetylated amino end; phosphoprotein
F;2-112/Product: cAMP-regulated phosphoprotein, 19K #status predicted <MAT>
F;17-112/Product: cAMP-regulated phosphoprotein, 16K #status predicted <MA2>
F;2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental
F;17/Modified site: acetylated amino end (Met) (in mature form) #status predicted

Query Match 70.3%; Score 425.5; DB 2; Length 112;
Best Local Similarity 76.4%; Pred. No. 1.2e-29;
Matches 84; Conservative 9; Mismatches 16; Indels 1; Gaps 1;

```
Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDS 67
           | |   : ||::: ::|    ||:||||||||:|| |||||||||| ||||||||||||
Db       4 EVPEAASAEEQKEMEDKV-TSPEKAEEAKLKARYPHLGQKPGGSDFLRKRLQKGQKYFDS 62

Qy      68 GDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           ||||||||||||||||:| |||  ||||||||||||||| ||| |||||
Db      63 GDYNMAKAKMKNKQLPTATPDKTEVTGDHIPTPQDLPQRKPSLVASKLAG 112
```

RESULT 2
JC4878
alpha endosulfine - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 30-Aug-1996 #sequence_revision 18-Oct-1996 #text_change 05-Nov-1999
C;Accession: JC4878
R;Peyrollier, K.; Heron, L.; Virsolvy-Vergine, A.; Le Cam, A.; Bataille, D.
Biochem. Biophys. Res. Commun. 223, 583-586, 1996
A;Title: Alpha endosulfine is a novel molecule, structurally related to a family of p
A;Reference number: JC4878; MUID:96268722
A;Contents: brain
A;Accession: JC4878

```
A;Molecule type: mRNA
A;Residues: 1-77 <PEY>
A;Cross-references: EMBL:X95958; NID:g1502303; PIDN:CAA65196.1; PID:e224652; PID:g150230
A;Experimental source: brain
C;Comment: This protein structurally related to a phosphoprotein family.


  Query Match           67.1%;  Score 406;  DB 2;  Length 77;
  Best Local Similarity  100.0%;  Pred. No. 3.7e-28;
  Matches   77;  Conservative    0;  Mismatches   0;  Indels    0;  Gaps    0;

Qy      25 EGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPS 84
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 EGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPS 60

Qy      85 AGPDKNLVTGDHIPTPQ 101
           |||||||||||||||||
Db      61 AGPDKNLVTGDHIPTPQ 77


RESULT   3
T23557
hypothetical protein K10C3.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T23557
R;Mortimore, B.
submitted to the EMBL Data Library, December 1996
A;Reference number: Z19759
A;Accession: T23557
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-174 <WIL>
A;Cross-references: EMBL:Z83235; PIDN:CAB05775.1; GSPDB:GN00019; CESP:K10C3.2
A;Experimental source: clone K10C3
C;Genetics:
A;Gene: CESP:K10C3.2
A;Map position: 1
A;Introns: 56/3; 132/3; 162/3


  Query Match           21.2%;  Score 128.5;  DB 2;  Length 174;
  Best Local Similarity  29.0%;  Pred. No. 0.00047;
  Matches   38;  Conservative   16;  Mismatches   24;  Indels   53;  Gaps    5;

Qy      29 PERAEEAKLKAKYPSLGQKPG--GSDFLMKRLQKGQKYFDSGDYNMAKAK----MKNKQL 82
           ||: :| :|  |  : |: |    | || |:||: :|:|||||| | |:|    : :|
Db      22 PEKQQEQELMGKLAATGKLPARPASSFLQKKLQQ-RKFFDSGDYAMDKSKAGTGLGSKPH 80

Qy      83 PSAG-------------------------------------------PDKNLVTGDHIPTP 100
           | ||                                           | ||     || |
Db      81 PLAGGPPPAAPPVVAQRSPAPAATTPSPSASPISQQTNRPSSDRNSDDDNL----QIPRP 136

Qy     101 QDLPQRKSSLV 111
           :||||:|::
Db     137 DTVPQRKASII 147


RESULT   4
T37510
hypothetical protein SPAC10F6.16 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T37510
R;Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, November 1997
A;Reference number: Z21719
A;Accession: T37510
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-139 <CON>
A;Cross-references: EMBL:AL009197; PIDN:CAA15729.1; GSPDB:GN00066; SPDB:SPAC10F6.16
A;Experimental source: strain 972h-; cosmid c10F6
C;Genetics:
A;Gene: SPDB:SPAC10F6.16
A;Map position: 1
A;Introns: 28/3; 56/3


  Query Match           20.2%;  Score 122;  DB 2;  Length 139;
  Best Local Similarity  34.1%;  Pred. No. 0.0013;
  Matches   30;  Conservative   15;  Mismatches   31;  Indels   12;  Gaps    3;

Qy      20 DTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKN 79
           ::::|  :    || ||   |   |: |   | |:::||:|:|||||||| : ||
Db      20 NSEQKVDVAKLSPEEQKLFRLY---GRLPQRKDLLVQKLQQGRKYFDSGDYALNKA---- 72

Qy      80 KQLPSAGPDKNLV-TGDHIPTPQDLPQR 106
             |  :   |  ||:|  :| |
Db      73 ----GKASDSGITCIGKEIPSPDTIPHR 96


RESULT   5
G90576
heat shock atp-dependent proteinase [imported] - Mycoplasma pulmonis (strain UAB CTIP
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: G90576
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer,
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: G90576
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-842 <KUR>
A;Cross-references: GB:AL445566; PID:g14089933; PIDN:CAC13692.1; GSPDB:GN00153
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU_5190
A;Genetic code: SGC3
C;Superfamily: ATP-dependent serine proteinase La


  Query Match           15.8%;  Score 95.5;  DB 2;  Length 842;
  Best Local Similarity  31.5%;  Pred. No. 1.7;
  Matches   35;  Conservative   19;  Mismatches   48;  Indels    9;  Gaps    4;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQ-KPGGSDFLMKRLQKG 61
           : |: ||  :|  ::    |:|| || |     |:||    ||:     ::  |   |
Db     234 EDQDLENAIDEKMKKNLSKQQKEFILRE-----KMKAIRSELGESNDDEEEYKTKIKDKN 288

Qy      62 QK--YFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSL 110
           |  |  |    :|  : | | : |: || | :| ::| |   || ||:|:
Db     289 NKKIYPKSVRKLIANEESKLKNMMSSSPDAN-ITRNYIDTLMKLPWRKTSI 338


RESULT   6
T28626
variant-specific surface protein 2 - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T28626
R;Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herrfeldt, J.A.; Peterson,
Cell 82, 89-100, 1995
A;Title: The large diverse gene family var encodes proteins involved in cytoadherence
A;Reference number: Z20487; MUID:95330813
A;Accession: T28626
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2664 <SUX>
A;Cross-references: EMBL:L40609; NID:g886376; PID:g886378; PIDN:AAA75398.1
C;Genetics:
A;Introns: 2197/3
A;Note: var-2
```

```
  Query Match             15.1%;  Score 91.5;  DB 2;  Length 2664;
  Best Local Similarity   26.8%;  Pred. No. 13;
  Matches   34;  Conservative   16;  Mismatches   30;  Indels   47;  Gaps    5;

Qy       6 EEENPAEETG-----EEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           ||||| |: |     ::: :  :| | | |: :| | : : |:  :: ||:
Db    2086 EEENPVEQPGFCPTPQQEPEPDDKCGKLEEKKDEKKEQPEQPA--EEDGGA--------- 2134

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGP------DKNLVTGDHIPTPQDLPQRK-----SS 109
                              :||  |      ||  |    || ||: |          |
Db    2135 --------------------IVPSGPPGSEPEADKGPVKPAEIPKPQEPPDLSHPAVIPS 2174

Qy     110 LVTSKLA 116
           |||| ||
Db    2175 LVTSTLA 2181


RESULT   7
S60970
hypothetical protein YNL157w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein N1743
C;Species: Saccharomyces cerevisiae
C;Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jul-2000
C;Accession: S60970; S63109; S63817
R;Nasr, F.; Becam, A.M.; Herbert, C.J.
submitted to the EMBL Data Library, October 1995
A;Description: The sequence of 36.8 kb from the left arm of chromosome XIV reveals 24 co
tonic dystrophy kinase.
A;Reference number: S60958
A;Accession: S60970
A;Molecule type: DNA
A;Residues: 1-168 <NAS>
A;Cross-references: EMBL:X92517; NID:g1050783; PID:g1050795
R;Nasr, F.; Becam, A.M.; Herbert, C.
submitted to the Protein Sequence Database, April 1996
A;Reference number: S62967
A;Accession: S63109
A;Molecule type: DNA
A;Residues: 1-168 <NAW>
A;Cross-references: EMBL:Z71433; NID:g1302119; PID:e239815; PID:g1302120; MIPS:YNL157w
A;Experimental source: strain S288C
R;Nasr, F.; Becam, A.M.; Herbert, C.J.
Yeast 12, 169-175, 1996
A;Title: The sequence of 36.8 kb from the left arm of chromosome XIV reveals 24 complete
dystrophy kinase.
A;Reference number: S63805; MUID:96287653
A;Accession: S63817
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-168 <NAF>
A;Cross-references: EMBL:X92517; NID:g1050783; PIDN:CAA63282.1; PID:g1050795
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995
C;Genetics:
A;Map position: 14L
A;Note: YNL157w


  Query Match             15.0%;  Score 90.5;  DB 2;  Length 168;
  Best Local Similarity   31.9%;  Pred. No. 0.8;
  Matches   30;  Conservative   11;  Mismatches   30;  Indels   23;  Gaps    5;

Qy      34 EAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKN--------KQLPSA 85
           | ||   |   |: |   |     :|| :|||||||| : || ::|          ||
Db      35 EMKL---YKMYGKLPSKKDIFKHTMQK-RKYFDSGDYALQKAGIQNNDPINYGKNNLPLT 90

Qy      86 GPDKNLVTGDHIPTPQDLPQRKSSLV--TSKLAG 117
             | |           :|: :|: |    |:  ||
Db      91 NPSK---------LREDIIKRRISTCPSTASTAG 115


RESULT   8
A84526
hypothetical protein At2g15190 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: A84526
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: A84526
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-775 <STO>
A;Cross-references: GB:AE002093; NID:g4585912; PIDN:AAD25573.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g15190
A;Map position: 2
C;Superfamily: Arabidopsis thaliana hypothetical protein T21L8.10


  Query Match             14.0%;  Score 85;  DB 2;  Length 775;
  Best Local Similarity   20.6%;  Pred. No. 12;
  Matches   32;  Conservative   24;  Mismatches   43;  Indels   56;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAE------EAKLKAKYPSLGQKPGGSDFLMK- 56
             :::||   |: | |:|   |:|| |  :||      |  ::      |  ||  :   :
Db     375 EEDEERKQEDEGVERQPEAEEEGGLERKAENDNESFEDSIREPNTQYGSYPGDDENTQRD 434

Qy      57 --------------------------------RLQKGQKYFDSGDYNMAKAKMKNKQL-P 83
                                            :|| :|       ::: : :||:: |
Db     435 VGDEVAEESSKDKSPTPRSSTPNFNILSEESLDVQKDKK-------RVSRGRNENKRVKP 487

Qy      84 SAGPDKNLVTGDHIPTPQ---------DLPQRKSS 109
           |   : || |   :|  :         |:| || |
Db     488 SVHAEDNLKTRKQVPRKRQKQVDSADVDVPTRKES 522


RESULT   9
EDBEQ3
immediate-early protein - saimiriine herpesvirus 1 (strain 11)
C;Species: saimiriine herpesvirus 1
A;Note: host Saimiri sciureus (common squirrel monkey)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
C;Accession: G36813; S20244
R;Albrecht, J.
submitted to the EMBL Data Library, January 1992
A;Description: Primary structure of the herpesvirus saimiri genome.
A;Reference number: A36806
A;Accession: G36813
A;Molecule type: DNA
A;Residues: 1-407 <ALB>
A;Cross-references: GB:X64346; NID:g60320; PIDN:CAA45696.1; PID:g60394
R;Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.; Newman, C.;
J. Virol. 66, 5047-5058, 1992
A;Title: Primary structure of the herpesvirus saimiri genome.
A;Reference number: A37309; MUID:92333688
A;Contents: annotation; possible protein-coding frames
A;Note: neither amino acid nor nucleotide sequence is given
R;Nicholas, J.; Cameron, K.R.; Honess, R.W.
Nature 355, 362-365, 1992
A;Title: Herpesvirus saimiri encodes homologues of G protein-coupled receptors and cy
A;Reference number: S20243; MUID:92115001
A;Accession: S20244
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-407 <NIC>
A;Cross-references: GB:S76368; NID:g243351; PIDN:AAB21116.1; PID:g243353
C;Genetics:
A;Gene: 73
C;Superfamily: herpesvirus immediate-early protein IE68
C;Keywords: early protein
```

```
  Query Match             13.8%;  Score 83.5;  DB 1;  Length 407;
  Best Local Similarity   30.6%;  Pred. No. 8.3;
  Matches   33;  Conservative   13;  Mismatches   35;  Indels   27;  Gaps    5;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAK----------YPSLGQK---- 47
           : :||   |||  || :: :| |    | ||||: : :           |  :|||
Db     203 EAEEEAEEAEEEAEEAEEAEEAEE-AEEEAEEAEEEEEEAGPSTPRLPHYKVVGQKPSTQ 261

Qy      48 PGGSDFL--------MKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGP 87
           |||   |          || ||::   | |    | : :||    | |
Db     262 PGGVPKLCLKMQPQHRSRLPKGKQ---SHDKVPKKYQARNKFFSQAAP 306


RESULT  10
A41971
surface protein pspA precursor - Streptococcus pneumoniae
N;Alternate names: pneumococcal surface protein A
C;Species: Streptococcus pneumoniae
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C;Accession: A41971; A60282; A33134
R;Yother, J.; Briles, D.E.
J. Bacteriol. 174, 601-609, 1992
A;Title: Structural properties and evolutionary relationships of PspA, a surface protein
A;Reference number: A41971; MUID:92105030
A;Accession: A41971
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-619 <YOT>
A;Cross-references: GB:M74122; NID:g153840; PIDN:AAA27018.1; PID:g153841
A;Note: sequence extracted from NCBI backbone (NCBIN:75635, NCBIP:75636)
R;Talkington, D.F.; Crimmins, D.L.; Voellinger, D.C.; Yother, J.; Briles, D.E.
Infect. Immun. 59, 1285-1289, 1991
A;Title: A 43-kilodalton pneumococcal surface protein, PspA: isolation, protective abili
A;Reference number: A60282; MUID:91169598
A;Accession: A60282
A;Molecule type: protein
A;Residues: 32-76 <TAL>
A;Experimental source: strain JY2008
C;Genetics:
A;Gene: pspA
C;Superfamily: cpl repeat homology
F;1-31/Domain: signal sequence #status predicted <SIG>
F;32-619/Product: surface protein pspA #status predicted <MAT>
F;411-430/Domain: cpl repeat homology <CP01>
F;431-450/Domain: cpl repeat homology <CP02>
F;451-470/Domain: cpl repeat homology <CP03>
F;471-490/Domain: cpl repeat homology <CP04>
F;491-510/Domain: cpl repeat homology <CP05>
F;511-530/Domain: cpl repeat homology <CP06>
F;531-550/Domain: cpl repeat homology <CP07>
F;551-570/Domain: cpl repeat homology <CP08>
F;571-591/Domain: cpl repeat homology <CP09>
F;592-611/Domain: cpl repeat homology <CP10>


  Query Match             13.7%;  Score 83;  DB 2;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 14;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  11
A97887
surface protein pspA precursor [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C;Accession: A97887
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: A97887
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-619 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAK98925.1; PID:g15457660; GSPDB:GN00174
C;Genetics:
A;Gene: pspA


  Query Match             13.7%;  Score 83;  DB 2;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 14;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  12
T08929
hypothetical protein T15N24.80 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C;Accession: T08929
R;Bevan, M.; Zimmermann, W.; Grueneisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.;
submitted to the Protein Sequence Database, May 1999
A;Reference number: Z16518
A;Accession: T08929
A;Molecule type: DNA
A;Residues: 1-763 <BEV>
A;Cross-references: EMBL:AL078465; GSPDB:GN00062; ATSP:T15N24.80
A;Experimental source: cultivar Columbia; BAC clone T15N24
C;Genetics:
A;Gene: ATSP:T15N24.80
A;Map position: 4
A;Introns: 345/3; 357/1; 384/3; 404/3; 441/3; 472/3; 500/3; 667/1; 697/3; 708/1


  Query Match             13.7%;  Score 83;  DB 2;  Length 763;
  Best Local Similarity   25.0%;  Pred. No. 18;
  Matches   31;  Conservative   25;  Mismatches   36;  Indels   32;  Gaps    6;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGI--LPERAEEAKLKAKYPSLGQKPGGSDFLMK---- 56
           :|:|||   ||  |||:: :|:|    :|:::|:   :|  ||  ::   |:   :
Db     526 EKEEEEKQEEEKAEEKEEKKEEENENGIPDKSED---EAPQPSESEEKDESEEHSEEETT 582

Qy      57 ------RLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSL 110
                 ||  |:|          :  : :||        | :|     | |: : |::||
Db     583 KKKRGSRLSAGKK--------ESAGRARNK--------KAVVAAKSSP-PEKITQKRSSA 625

Qy     111 VTSK 114
               |
Db     626 KRKK 629


RESULT  13
E84525
hypothetical protein At2g15140 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
```

```
C;Accession: E84525
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: E84525
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-792 <STO>
A;Cross-references: GB:AE002093; NID:g4585915; PIDN:AAD25576.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g15140
A;Map position: 2
C;Superfamily: Arabidopsis thaliana hypothetical protein T21L8.10


  Query Match             13.7%;  Score 83;  DB 2;  Length 792;
  Best Local Similarity   21.3%;  Pred. No. 19;
  Matches   33;  Conservative   25;  Mismatches   41;  Indels   56;  Gaps    6;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAE------EAKLKAKYPSLGQKPGG------- 50
           :::||   |: | |:|   |:|| |  :||     |  ::      |  ||
Db     386 EEDEERKQEDEGVERQPEAEEEGGLERKAENDNESFEDSIREPNTQYGSYPGDDENTQRD 445

Qy      51 -SDFLMKR------------------------LQKGQKYFDSGDYNMAKAKMKNKQL-P 83
            | |::                         :|| :|       ::: : :||:: |
Db     446 VGDELVEESSKDKSPTPRSSTPNFNILSEESLDVQKDKK-------RVSRGRNQNKRVKP 498

Qy      84 SAGPDKNLVTGDHIPTPQ---------DLPQRKSS 109
           |   : || |   :|  :          |:| || :
Db     499 SVHAEDNLKTRKQVPRKRQKQVDSADVDVPTRKEA 533


RESULT  14
D84477
hypothetical protein At2g06430 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: D84477
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: D84477
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-795 <STO>
A;Cross-references: GB:AE002093; NID:g6598730; PIDN:AAD26891.2; GSPDB:GN00139
C;Genetics:
A;Gene: At2g06430
A;Map position: 2
C;Superfamily: Arabidopsis thaliana hypothetical protein T21L8.10


  Query Match             13.6%;  Score 82.5;  DB 2;  Length 795;
  Best Local Similarity   25.4%;  Pred. No. 21;
  Matches   33;  Conservative   22;  Mismatches   38;  Indels   37;  Gaps    6;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKR---LQK 60
           :::||   |: | |:|   |:|| |  :||:     | |:         : | :    :||
Db     405 EEDEERKQEDEGVERQPEAEEEGGLERKAED-----KSPTPRSTTPNFNILSEESLDVQK 459

Qy      61 GQKYFDSGDYNMAKAKMKNKQL-PSAGPDKNLVT---------------GDHIPTP---- 100
            :|        ::: : :||:: ||   : || |                ||:
Db     460 DKK-------RVSRGRNENKRVKPSVHAEDNLKTRKQAQSKKRKIIGNDGDNADNDGDND 512

Qy     101 --QDLPQRKS 108
             |  |||||
Db     513 DFQPAPQRKS 522


RESULT  15
S55395
neurofilament protein M - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 13-Aug-1999
C;Accession: S55395
R;Vitadello, M.; Vettore, S.; Lamar, E.; Chien, K.R.; Gorza, L.
submitted to the EMBL Data Library, January 1995
A;Description: Neurofilament mRNA and protein are expressed in precursors of heart co
A;Reference number: S55395
A;Accession: S55395
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-644 <VIT>
A;Cross-references: EMBL:Z47378; NID:g854352; PIDN:CAA87454.1; PID:g854353
C;Superfamily: cytoskeletal keratin


  Query Match             13.6%;  Score 82;  DB 2;  Length 644;
  Best Local Similarity   28.9%;  Pred. No. 18;
  Matches   22;  Conservative   17;  Mismatches   27;  Indels   10;  Gaps    2;

Qy       5 QEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKY 64
           :|||   || |:|::: :| ||:  ::|||          | :  ||       ::|:
Db     308 KEEEGEKEEEGQEEEEEEEDEGVKSDQAEEG---------GSEKEGSSKNEGEQEEGETE 358

Qy      65 FDSGDYNMAKAKMKNK 80
            : |:   |:|| : |
Db     359 AE-GEVEEAEAKEEKK 373


RESULT  16
S25776
transforming protein (SHC) - human
C;Species: Homo sapiens (man)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C;Accession: S25776
R;Pelicci, G.; Lanfrancone, L.; Grignani, F.; McGlade, J.; Cavallo, F.; Forni, G.; Ni
Cell 70, 93-104, 1992
A;Title: A novel transforming protein (SHC) with an SH2 domain is implicated in mitog
A;Reference number: S25776; MUID:92323554
A;Accession: S25776
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-473 <PEL>
A;Cross-references: EMBL:X68148; NID:g36453; PIDN:CAA48251.1; PID:g36454
C;Superfamily: SH2 homology
F;378-467/Domain: SH2 homology <SH2B>


  Query Match             13.5%;  Score 81.5;  DB 2;  Length 473;
  Best Local Similarity   26.2%;  Pred. No. 15;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   33;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|       | : :||  |  |      |: ||       :||
Db     229 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 288

Qy      47 KPGGSDFLMKRLQ-----KGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDHI 97
             ||   : |::       |:: ||   |    |: ||:   : :  |||    : |
Db     289 PVGGDPEVRKQMPPPPPCPGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS-- 343

Qy      98 PTPQDL 103
             |:||
Db     344 -APRDL 348


RESULT  17
T02398
hypothetical protein At2g44430 [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein F4I1.24
```

```
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C;Accession: T02398; D84878
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, May 1998
A;Description: Arabidopsis thaliana chromosome II BAC F4I1 genomic sequence.
A;Reference number: Z14667
A;Accession: T02398
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-646 <ROU>
A;Cross-references: EMBL:AC004521; NID:g3128166; PID:g3128185
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: D84878
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-646 <STO>
A;Cross-references: GB:AE002093; NID:g3128185; PIDN:AAC16089.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g44430; F4I1.24
A;Map position: 2
A;Introns: 156/2; 335/3

  Query Match             13.5%;  Score 81.5;  DB 2;  Length 646;
  Best Local Similarity   25.8%;  Pred. No. 20;
  Matches   32;  Conservative   17;  Mismatches   52;  Indels   23;  Gaps    4;

Qy       3 QKQ-EEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKG 61
           ||| | :  :      ||||             | : |  |  :|   :||| |||:|
Db     515 QKQTESKTNSSNDNSSKQDT------------GKTEKKTVSADKKKSVADFL-KRLKKN 560

Qy      62 QKYFDSGDYNM--------AKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTS 113
             :: | |          :| | :  :  | |  |  |    :      ||:|::  |:
Db     561 SPQKEAKDQNKSGGNVKKDSKTKPRELRSSSVGKKKAEVENTPVKRAPGRPQKKTAEATA 620

Qy     114 KLAG 117
           :|
Db     621 SASG 624


RESULT  18
E84514
hypothetical protein At2g14130 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: E84514
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: E84514
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-789 <STO>
A;Cross-references: GB:AE002093; NID:g4587682; PIDN:AAD25852.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g14130
A;Map position: 2
C;Superfamily: Arabidopsis thaliana hypothetical protein T21L8.10

  Query Match             13.5%;  Score 81.5;  DB 2;  Length 789;
  Best Local Similarity   19.3%;  Pred. No. 25;
  Matches   28;  Conservative   28;  Mismatches   50;  Indels   39;  Gaps    4;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAE------EAKLKAKYPSLGQKPGGSDFLMK- 56
           :::||   |: | |:|   |:|| |  :||      |  ::      |  ||  :   :
Db     386 EEDEERKQEDEGVERQPEAEEEGGLERKAENDNESFEDSIREPNTQYGSYPGDDENTQRD 445

Qy      57 ----------------------RLQKGQKYFDSGDYNMAKAKMKNKQL-PSAGPDKNLVT 93
                                    : ::     : :::  :::||:: ||   : || |
Db     446 VGDEVAEESSKDKSHTPRSSTPNFNRIRRELAEEEMSLSFMQLQNKRVKPSVHAEDNLKT 505

Qy      94 GDHIPTPQ---------DLPQRKSS 109
              :|  :          |:| || :
Db     506 RKQVPRKRQKQVDSADVDVPTRKEA 530


RESULT  19
JQ0797
histone H2B.IV - Volvox carteri
C;Species: Volvox carteri
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 23-Jul-1999
C;Accession: JQ0797
R;Mueller, K.; Lindauer, A.; Bruederlein, M.; Schmitt, R.
Gene 93, 167-175, 1990
A;Title: Organization and transcription of Volvox histone-encoding genes: similaritie
A;Reference number: JQ0794; MUID:91033024
A;Accession: JQ0797
A;Molecule type: DNA
A;Residues: 1-155 <MUE>
A;Cross-references: GB:M31922; NID:g170657; PIDN:AAA34250.1; PID:g170659
C;Genetics:
A;Gene: H2B-IV
C;Superfamily: histone H2B
C;Keywords: chromosomal protein; DNA binding; nucleosome core; nucleus

  Query Match             13.4%;  Score 81;  DB 2;  Length 155;
  Best Local Similarity   29.2%;  Pred. No. 4.8;
  Matches   26;  Conservative   13;  Mismatches   42;  Indels    8;  Gaps    2;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEE---AKLKAKYPSLGQKPGGSDFLMKR 57
           |: | :|| || |  | | : : |    |:: |:   ||  || |: |    |      :
Db       1 MAPKTKEEKPASEAVEPKAEAKPKAEKAPKKKEKKAPAKKSAKEPAAGDAAEGD-----K 55

Qy      58 LQKGQKYFDSGDYNMAKAKMKNKQLPSAG 86
           :|  |   |  | :   |:  :  |  |
Db      56 KKKKAKVAKSETYKLYIYKVLKQVHPDTG 84


RESULT  20
I38616
zinc finger protein ZNF139 - human (fragment)
N;Alternate names: finger protein kox18
C;Species: Homo sapiens (man)
C;Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 01-Dec-2000
C;Accession: I38616; I37957; S10406
R;Tommerup, N.; Vissing, H.
Genomics 27, 259-264, 1995
A;Title: Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs iden
A;Reference number: A57785; MUID:96044430
A;Accession: I38616
A;Molecule type: mRNA
A;Residues: 1-325 <RES>
A;Cross-references: EMBL:U09848; NID:g495567; PIDN:AAC50264.1; PID:g495568
R;Thiesen, H.J.
New Biol. 2, 363-374, 1990
A;Title: Multiple genes encoding zinc finger domains are expressed in human T cells.
A;Reference number: I37949; MUID:91145339
A;Accession: I37957
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 167-222 <RE2>
A;Cross-references: EMBL:X52349; NID:g34150; PIDN:CAA36575.1; PID:g930080
C;Genetics:
A;Gene: GDB:ZNF139
```

```
A;Cross-references: GDB:137044
A;Map position: 7q21.3-7q22.1
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: DNA binding; zinc finger


  Query Match             13.4%;  Score 81;  DB 2;  Length 325;
  Best Local Similarity   21.3%;  Pred. No. 11;
  Matches   32;  Conservative   25;  Mismatches   47;  Indels   46;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
Db      18 RQENYGSAFPQGGENRNENEESTSKAETSEDSASRGETTGRSQK----EFGEKRDQEGK- 72

Qy      64 YFDSGDYNMAKAKMKNKQL-----PSAGPDKNLVTGDHIP----------------TP 100
Db      73 ---TGERQQKNPEEKTRKEKRDSGPAIGKDKKTITGERGPREKGKGLGRSFSLSSNFTTP 129

Qy     101 QDLP--------------QRKSSLVTSKL 115
Db     130 EEVPTGTKSHRCDECGKCFTRSSSLIRHKI 159


RESULT  21
I51652
dsRNA-binding protein 4F.1 - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: I51652
R;Bass, B.L.; Hurst, S.R.; Singer, J.D.
Curr. Biol. 4, 301-314, 1994
A;Title: Binding properties of newly identified Xenopus proteins containing dsRNA-bindi
A;Reference number: I51652; MUID:95006648
A;Accession: I51652
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-695 <BAS>
A;Cross-references: EMBL:U07155; NID:g463902; PIDN:AAA19960.1; PID:g514259
C;Superfamily: double-stranded RNA-binding repeat homology
F;401-469/Domain: double-stranded RNA-binding repeat homology <DSR1>


  Query Match             13.4%;  Score 81;  DB 2;  Length 695;
  Best Local Similarity   21.6%;  Pred. No. 24;
  Matches   36;  Conservative   19;  Mismatches   48;  Indels   64;  Gaps    7;

Qy       1 MSQKQE-----EENPAEETGEEKQDTQEKEGILPERAEEAKLKA---------------- 39
Db     472 MEEKEEGTDESEQKPVVQTPAQPDDSAEVDSAALDQAESAKQQGPILTKHGKNPVMELNE 531

Qy      40 -----KYPSLGQKPGGSD--FLMKRLQKGQKYFDSG-------------------DYNM 72
Db     532 KRRGLKYEMISETGGSHDKRFVMEVEVDGVKFQGSGSNKKVAKAYAALSALEKLFPDYTM 591

Qy      73 ---AKAKMKNKQLPSAGPDKNLVTGDH----------IPTPQDLPQR 106
Db     592 YTEAPKKKRPPMMPRGGPK---FAGKHNQGFGMMYSEVPPPQAMRGR 635


RESULT  22
C96608
hypothetical protein F25P12.91 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: C96608
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: C96608
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-522 <STO>
A;Cross-references: GB:AE005173; NID:g9954748; PIDN:AAG09099.1; GSPDB:GN00141
C;Genetics:
A;Gene: F25P12.91
A;Map position: 1


  Query Match             13.3%;  Score 80.5;  DB 2;  Length 522;
  Best Local Similarity   19.7%;  Pred. No. 20;
  Matches   25;  Conservative   27;  Mismatches   56;  Indels   19;  Gaps    2;

Qy       3 QKQEEENPAEETGEEKQDTQEKEG------ILPERAEEAKLKAKYPSLGQKPGGSDFLMK 56
Db      43 EKVKAKKDEESSGKSKKDKEKKKGKNVDSEVKEDKDDDKKKDGKMVSKKHEEGHGDLEVK 102

Qy      57 ------------RLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDL 103
Db     103 ESDVKVEEHEKEHKKGKEKKHEELEEEKEGKKKKNKKEKDESGPEEKNKKADKEKKHEDV 162

Qy     104 PQRKSSL 110
Db     163 SQEKEEL 169


RESULT  23
T12972
hypothetical protein T21L8.10 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 18-Aug-2000
C;Accession: T12972
R;Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.;
submitted to the Protein Sequence Database, July 1999
A;Reference number: Z17586
A;Accession: T12972
A;Molecule type: DNA
A;Residues: 1-820 <CHO>
A;Cross-references: EMBL:AL096860; GSPDB:GN00061; ATSP:T21L8.10
A;Experimental source: cultivar Columbia; BAC clone T21L8
C;Genetics:
A;Gene: ATSP:T21L8.10
A;Map position: 3
A;Introns: 115/1; 297/3; 431/3; 443/3; 484/3; 520/3; 539/3; 560/3; 641/2; 668/3; 726/
C;Superfamily: Arabidopsis thaliana hypothetical protein T21L8.10


  Query Match             13.3%;  Score 80.5;  DB 2;  Length 820;
  Best Local Similarity   20.1%;  Pred. No. 32;
  Matches   28;  Conservative   28;  Mismatches   52;  Indels   31;  Gaps    4;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAE------EAKLKAKYPSLGQKPGGSDFLMKR 57
Db     414 EEDEERKQEDEGVERQPEAEEEGGLERKAENDNESFEDSIREPNTQYGSYPGDDENIQRD 473

Qy      58 LQKGQKYFDSGDYNMAKAK----------------MKNKQLPSAGPDKN------LVTG 94
Db     474 V--GDELVEESSKDMSPTPRSSTLNFNILSEESLDVQKDKKRVSRGRNENKRVKPSVHAE 531

Qy      95 DHIPTPQDLPQRKSSLVTS 113
Db     532 DNFKTRKQMPRKRQKQVDS 550


RESULT  24
T23482
hypothetical protein K08F4.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
```

```
C;Accession: T23482
R;Hembry, C.
submitted to the EMBL Data Library, January 1996
A;Reference number: Z19746
A;Accession: T23482
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-215 <WIL>
A;Cross-references: EMBL:Z68879; PIDN:CAA93085.1; GSPDB:GN00022; CESP:K08F4.5
A;Experimental source: clone K08F4
C;Genetics:
A;Gene: CESP:K08F4.5
A;Map position: 4
A;Introns: 22/1


  Query Match             13.2%;  Score 80;  DB 2;  Length 215;
  Best Local Similarity   27.2%;  Pred. No. 8.3;
  Matches   34;  Conservative   21;  Mismatches   40;  Indels   30;  Gaps    6;

Qy       3 QKQEEENPA---EETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQ 59
           :|:||:  |   ||  |||:: : ||    |: ||:|   |  |  :|    |   :: :
Db      79 KKEEEKKDATKKEEAKEEKKEEERKE----EKKEESKEGEKKNSGEEKKNSGD---EKKE 131

Qy      60 KGQKYFDSGD-YNMAKAKMKNKQLPSAGPDKNLV-----------TGDHI-------PT 99
           :| |  |: |     :||  :|:     | | :|              | :        ||
Db     132 EGGKKDDAKDGEKKEEAKEGDKKDEDQKPKKEIVPEVFKRTSKEKAIADGVKKKKTDYPT 191

Qy     100 PQDLP 104
           :|:|
Db     192 MEDVP 196


RESULT  25
S65469
DNA topoisomerase (EC 5.99.1.2) - Caenorhabditis elegans
N;Alternate names: DNA topoisomerase I
C;Species: Caenorhabditis elegans
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-May-1997
C;Accession: S65469
R;Kim, J.; Kim, Y.C.; Lee, J.H.; Jang, Y.J.; Chung, I.K.; Koo, H.S.
Eur. J. Biochem. 237, 367-372, 1996
A;Title: cDNA cloning, expression, and chromosomal localization of Caenorhabditis elegan
A;Reference number: S65469; MUID:96215431
A;Accession: S65469
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-751 <KIM>
A;Cross-references: EMBL:X96762
C;Genetics:
A;Gene: topoI
C;Superfamily: eukaryotic type I DNA topoisomerase
C;Keywords: isomerase


  Query Match             13.2%;  Score 80;  DB 2;  Length 751;
  Best Local Similarity   32.6%;  Pred. No. 32;
  Matches   30;  Conservative   12;  Mismatches   26;  Indels   24;  Gaps    4;

Qy       6 EEENPAEET-GEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKY 64
           |||: |  |  ||:|  :|||    |: |: | | |              |||:|  :
Db      86 EEEDEAPTTDDEEEQKRKEKERRKKEKREKKKRKEK--------------KRLEKENRK 130

Qy      65 F------DSGDYNMAKAKMKNKQLPSAGPDKN 90
                 || | :  ||| | ::  | | :|:
Db     131 IKEEDDEDSDDEDDEKAKKKKRK--SKGAEKS 160


RESULT  26
T33457
hypothetical protein F36H12.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33457
R;Blanchard, M.; Bradshaw, H.; Stellyes, L.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of C. elegans cosmid F36H12.
A;Reference number: Z21346
A;Accession: T33457
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-335 <BLA>
A;Cross-references: EMBL:AF078790; PIDN:AAC26930.1; GSPDB:GN00022; CESP:F36H12.3
A;Experimental source: strain Bristol N2; clone F36H12
C;Genetics:
A;Gene: CESP:F36H12.3
A;Map position: 4
A;Introns: 32/3; 227/1; 270/2


  Query Match             13.1%;  Score 79;  DB 2;  Length 335;
  Best Local Similarity   29.0%;  Pred. No. 16;
  Matches   36;  Conservative   20;  Mismatches   52;  Indels   16;  Gaps    6;

Qy       3 QKQEE--ENPAEETGEEKQDTQEKEGILPERAE-EAKLKAK----YPSLGQKPGGSDFLM 55
           :|:||  |   ||  |||:: ::||    |: | | |:: |     | :  :| | | :
Db     179 EKKEEKKEEKKEEKKEEKKEEEKKEEKKEEKKEKEDEKKVEPKPGEPKPHISIEPPG-DLMF 237

Qy      56 KRLQKGQKYFDSGDYN----MAKAKMKNKQLPSAGPDKNLV----TGDHIPTPQDLPQRK 107
           |  |: ||      : :    | | |  : |:    |    |    | |   |    | ::
Db     238 KADQQEQKKLTLTNTHDKKIMFKIKTSDNQVYLVNPVYGTVEPGKTADLTITRNKAPAKE 297

Qy     108 SSLV 111
           : ||
Db     298 AKLV 301


RESULT  27
C88131
protein F10G7.1 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: C88131
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A;Accession: C88131
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-785 <STO>
A;Cross-references: GB:chr_II; PIDN:AA81131.1; PID:g1055162; GSPDB:GN00020; CESP:F10G
C;Genetics:
A;Gene: F10G7.1
A;Map position: 2


  Query Match             13.1%;  Score 79;  DB 2;  Length 785;
  Best Local Similarity   23.1%;  Pred. No. 41;
  Matches   24;  Conservative   16;  Mismatches   44;  Indels   20;  Gaps    3;

Qy       6 EEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           :|::  ||  :|  |  : | :  :  ||  :       |:                : |
Db     378 DEDDEDEEDSDEDMDDSDNEEVEDDSEEEEPMDDLKSEAGETTA----------SEMMF 426

Qy      66 DSG---DYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQR 106
           | |   | |||: :   |:       :|    | : || |:| |
Db     427 DDGIDEDINMAEVEKYRKE------RENAQWPDEVDTPMDMPAR 464


RESULT  28
S15762
neurofilament triplet M protein - chicken
C;Species: Gallus gallus (chicken)
```

```
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 13-Aug-1999
C;Accession: S15762; S08061; A27040
R;Zopf, D.; Dineva, B.; Betz, H.; Gundelfinger, E.D.
Nucleic Acids Res. 18, 521-529, 1990
A;Title: Isolation of the chicken middle-molecular weight neurofilament (NF-M) gene and
A;Reference number: S15762; MUID:90174973
A;Accession: S15762
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-858 <ZOP>
A;Cross-references: EMBL:X17102; NID:g63688; PIDN:CAA34958.1; PID:g63689
R;Zopf, D.; Dineva, B.; Betz, H.; Gundelfinger, E.
submitted to the EMBL Data Library, November 1989
A;Reference number: S08061
A;Accession: S08061
A;Molecule type: DNA
A;Residues: 1-355,368-858 <ZO2>
A;Cross-references: EMBL:X17102
R;Zopf, D.; Hermans-Borgmeyer, I.; Gundelfinger, E.D.; Betz, H.
Genes Dev. 1, 699-708, 1987
A;Title: Identification of gene products expressed in the developing chick visual system
A;Reference number: A27040; MUID:88112814
A;Accession: A27040
A;Molecule type: mRNA
A;Residues: 350-546,'R',548-858 <ZO3>
A;Cross-references: GB:X05558; NID:g63685; PIDN:CAA29073.1; PID:g63686
C;Genetics:
A;Introns: 355/3; 397/2
C;Superfamily: cytoskeletal keratin
C;Keywords: coiled coil


  Query Match             13.1%;  Score 79;  DB 2;  Length 858;
  Best Local Similarity   25.2%;  Pred. No. 45;
  Matches   36;  Conservative   18;  Mismatches   59;  Indels   30;  Gaps    5;

Qy       3 QKQEEE---------NPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQK------ 47
           :|:|||         : ||| | :|:: :|||       |||: | |    | |
Db     514 EKEEEEAEEEEAAKSDAAEEGGSKKEEIEEKEEGEEAEEEEAEAKGKAEEAGAKVEKVKS 573

Qy      48 ---------PGGSDFL--MKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNL---VT 93
                  |  |      | :||           |:   |:  :| |:|     ||
Db     574 PPAKSPPKSPPKSPVTEQAKAVQKAAAEVGKDQKAEKAAEKAAKEEKAASPEKPATPKVT 633

Qy      94 GDHIP-TPQDLPQRKSSLVTSKL 115
               | ||:  |  : ::   |:
Db     634 SPEKPATPEKPPTPEKAITPEKV 656


RESULT  29
T31433
Na+/Ca2+,K+-exchanging protein - bottle-nosed dolphin
C;Species: Tursiops truncatus (bottle-nosed dolphin)
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T31433
R;Cooper, C.B.; Winkfein, R.J.; Schnetkamp, P.P.M.
submitted to the EMBL Data Library, April 1998
A;Description: Cloning, sequencing, and functional expression of the bottlenose dolphin
A;Reference number: Z21032
A;Accession: T31433
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1014 <COO>
A;Cross-references: EMBL:AF059031; NID:g3170602; PID:g3170603; PIDN:AAC18119.1
C;Genetics:
A;Gene: NCKXr


  Query Match             13.1%;  Score 79;  DB 2;  Length 1014;
  Best Local Similarity   29.3%;  Pred. No. 54;
  Matches   22;  Conservative   14;  Mismatches   25;  Indels   14;  Gaps    2;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLM-------- 55
           : |:|   || |||::: :|:|    |  ||  |  ::|   ||     ||:
Db     779 ESEDEEEEEEEGEEEEQEEEEEEQEEEEEEEPLSLEWPETRQKQAIYLFLLPIVFPLWL 838

Qy      56 -----KRLQKGQKYF 65
                 :||:  :|:|
Db     839 TVPDVRRLE-ARKFF 852


RESULT  30
S72269
ryanodine receptor isoform 2, cardiac muscle - human
N;Alternate names: calcium release channel protein isoform 2
C;Species: Homo sapiens (man)
C;Date: 14-Apr-1998 #sequence_revision 24-Apr-1998 #text_change 20-Jun-2000
C;Accession: S72269
R;Tunwell, R.E.A.; Wickenden, C.; Bertrand, B.M.A.; Shevchenko, V.I.; Walsh, M.B.; Al
Biochem. J. 318, 477-487, 1996
A;Title: The human cardiac muscle ryanodine receptor-calcium release channel: identif
A;Reference number: S72269; MUID:96404895
A;Accession: S72269
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-4967 <TUN>
A;Cross-references: EMBL:X98330; NID:g1526977; PIDN:CAA66975.1; PID:g1526978
A;Experimental source: cardiac muscle; sarcoplasmic reticulum
C;Genetics:
A;Gene: RyR-2
C;Complex: homotetramer
C;Function:
A;Description: mediates rapid release of Ca2+ ions from intracellular stores during c
C;Superfamily: ryanodine receptor; transcription initiation factor sigma region 1 hom
C;Keywords: alternative splicing; calcium channel; cardiac muscle; glycoprotein; hear
F;710-715/Region: adenine nucleotide binding #status predicted
F;1148-1153/Region: adenine nucleotide binding #status predicted
F;2338-2342/Region: adenine nucleotide binding #status predicted
F;2626-2631/Region: adenine nucleotide binding #status predicted
F;4030-4048/Domain: calcium binding #status predicted <CA1>
F;4499-4519/Domain: transmembrane #status predicted <TM01>
F;4572-4593/Domain: transmembrane #status predicted <TM02>
F;4719-4744/Domain: transmembrane #status predicted <TM03>
F;4767-4789/Domain: transmembrane #status predicted <TM04>
F;4809-4829/Domain: transmembrane #status predicted <TM05>
F;4844-4867/Domain: transmembrane #status predicted <TM06>
F;140,301/Binding site: phosphate (Thr) (covalent) (by cAMP-and calmodulin-dependent
F;2808/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #sta
F;4794/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             13.1%;  Score 79;  DB 2;  Length 4967;
  Best Local Similarity   23.9%;  Pred. No. 3e+02;
  Matches   28;  Conservative   17;  Mismatches   42;  Indels   30;  Gaps    3;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL-----GQKPGGSDFLMKRL 58
           |:||:   |||  | :  : ::|   |:|:| | | |   |     |:         |::
Db    4425 KEEEKEEKEETKSEPEKAEGEDGEKEEKAKEDKGKQKLRQLHTHRYGEPEVPESAFWKKI 4484

Qy      59 ----QKGQKYFDSGDYNM--------------------AKAKMKNKQLPSAGPDKN 90
               ||   ||    |||                      : :: |:||:    :|
Db    4485 IAYQQKLLNYFARNFYNMRMLALFVAFAINFILLFYKVSTSSVVEGKELPTRSSSEN 4541


RESULT  31
S60480
low temperature-induced protein lti29 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 01-Dec-2000
C;Accession: S60480; S52655; S49029
R;Welin, B.V.; Olson, A.; Palva, E.T.
Plant Mol. Biol. 29, 391-395, 1995
A;Title: Structure and organization of two closely related low-temperature-induced dh
A;Reference number: S60480; MUID:96046757
A;Accession: S60480
A;Molecule type: DNA
```

```
A;Residues: 1-260 <WEL>
A;Cross-references: EMBL:X90958; NID:g975647; PIDN:CAA62448.1; PID:g975648
R;Welin, B.V.; Olson, A.; Nylander, M.; Palva, E.T.
Plant Mol. Biol. 26, 131-144, 1994
A;Title: Characterization and differential expression of dhn/lea/rab-like genes during c
A;Reference number: S52654; MUID:95035985
A;Accession: S52655
A;Molecule type: mRNA
A;Residues: 'G',57-143,'E',145-260 <WE2>
A;Cross-references: EMBL:X77614; NID:g509261; PIDN:CAA54705.1; PID:g509262
A;Experimental source: ecotype Landsberg erecta
C;Genetics:
A;Gene: lti29; lti45
A;Introns: 110/3


  Query Match             13.0%;  Score 78.5;  DB 2;  Length 260;
  Best Local Similarity   28.4%;  Pred. No. 14;
  Matches   42;  Conservative   19;  Mismatches   42;  Indels   45;  Gaps   11;

Qy       2 SQKQEEENPAEETGEEKQDTQEK---EGILPERAEEA------KLKAKYPSLGQKPGGSD 52
           |   :||    | ||:|:  ::|   ||   :  ||       ::| |:| ||:|||| |
Db     109 SSSSDEEG---EDGEKKKKEKKKKIVEGDHVKTVEEENQGVMDRIKEKFP-LGEKPGGDD 164

Qy      53 FLM------------KRLQKGQKYFDSGDYNMAKAKMKNKQLP--SAGPDKNLVTGD-- 95
             :             | :: :| |      | | | |   ||  |  |: : |
Db     165 VPVVTTMPAPHSVEDHKPEEEEKKGF------MDKIKEK---LPGHSKKPEDSQVVNTTP 215

Qy      96 --HIPTP-QDLPQRKSSL---VTSKLAG 117
               ||  |:|: |      :  || |
Db     216 LVETATPIADIPEEKKGFMDKIKEKLPG 243


RESULT  32
T34036
hypothetical protein B0041.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34036
R;Fulton, R.; Wohldmann, P.
submitted to the EMBL Data Library, April 1997
A;Description: The sequence of C. elegans cosmid B0041.
A;Reference number: Z21466
A;Accession: T34036
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1359 <FUL>
A;Cross-references: EMBL:AF000196; PIDN:AAC24256.1; GSPDB:GN00019; CESP:B0041.7
A;Experimental source: strain Bristol N2; clone B0041
C;Genetics:
A;Gene: CESP:B0041.7
A;Map position: 1
A;Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3


  Query Match             13.0%;  Score 78.5;  DB 2;  Length 1359;
  Best Local Similarity   26.9%;  Pred. No. 82;
  Matches   32;  Conservative   24;  Mismatches   48;  Indels   15;  Gaps    5;

Qy       2 SQKQEEENPAEETGEE---KQDTQEKEGILPERAE--EAKLKAKYPSLGQKPGGSDFLMK 56
           ::||     :||: ||   |:  : ||  : :|||  |   : : ||   | |     :|
Db     147 TKKQTSSESSEESEEERKVKKSKKNKEKSVKKRAETSEESDEDEKPSKKSKKG-----LK 201

Qy      57 RLQKGQKYFDSGD-YNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSK 114
           :  | :   :| |   : |:| |:|::      |   : |  |  :   :|| |  :|:
Db     202 KKAKSESESESEDEKEVKKSKKKSKKVVK----KESESEDEAPEKKKTEKRKRSKTSSE 256


RESULT  33
T41384
hypothetical protein SPCC550.09 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000
C;Accession: T41384
R;Murphy, L.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, May 1998
A;Reference number: Z21990
A;Accession: T41384
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-535 <MUR>
A;Cross-references: EMBL:AL023592; PIDN:CAA19113.1; GSPDB:GN00068; SPDB:SPCC550.09
A;Experimental source: strain 972h-; cosmid c550
C;Genetics:
A;Gene: SPDB:SPCC550.09
A;Map position: 3
C;Superfamily: Schizosaccharomyces pombe hypothetical protein SPCC550.09


  Query Match             12.9%;  Score 78;  DB 2;  Length 535;
  Best Local Similarity   23.8%;  Pred. No. 33;
  Matches   20;  Conservative   19;  Mismatches   45;  Indels    0;  Gaps    0;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
             : ::|  ::|: |   |    :  |   |   || :  |:     ::   |  : :  :|
Db     150 VKEQQASQSPSVEGENEPAGVKSNEATKPSVTEEKEKNAEQNKRRERLSISHIVTELAKK 209

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPS 84
             :   :||   :  ||:||:|| |
Db     210 DEVSSESGSNEIVNAKLKNEQLDS 233


RESULT  34
T48930
hypothetical protein F14L2.50 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Sep-2000
C;Accession: T48930
R;Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z25008
A;Accession: T48930
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-742 <JOR>
A;Cross-references: EMBL:AL353818; GSPDB:GN00061; ATSP:F14L2.50
A;Experimental source: cultivar Columbia; BAC clone F14L2
C;Genetics:
A;Gene: ATSP:F14L2.50
A;Map position: 3
A;Introns: 115/1; 211/3; 336/3; 348/3; 389/3; 425/3; 444/3; 465/3; 546/2; 573/3; 607/
C;Superfamily: Arabidopsis thaliana hypothetical protein T21L8.10


  Query Match             12.9%;  Score 78;  DB 2;  Length 742;
  Best Local Similarity   22.6%;  Pred. No. 47;
  Matches   24;  Conservative   22;  Mismatches   44;  Indels   16;  Gaps    3;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAE------EAKLKAKYPSLGQKPGGSDFLMKR 57
             :::||   |: | |:|   |:|| |  :||      |  ::      |  ||  :   :
Db     319 EEDEERKQEDEGVERQPEAEEEGGLERKAENDNESFEDSIREPNTQYGSYPGDDENTQR- 377

Qy      58 LQKGQKYFDSGDYNMAKAKMKNKQLP-SAGPDKNLVTGDHIPTPQD 102
                 | ||  : ::       | |: |: |::: : :   :|
Db     378 --------DVGDELVEESSKDKSPTPRSSTPNFNILSEESLDVQKD 415


RESULT  35
H83736
subtilisin-type alkaline proteinase (EC 3.4.21.-) apr precursor [similarity] - Bacill
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: H83736
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
```

```
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: H83736
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-792 <STO>
A;Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BAB04415.1; GSPDB:GN0
A;Experimental source: strain C-125
C;Genetics:
A;Gene: apr
C;Keywords: hydrolase; serine proteinase
F;1-28/Domain: signal sequence #status predicted <SIG>


  Query Match             12.9%;  Score 78;  DB 2;  Length 792;
  Best Local Similarity   26.5%;  Pred. No. 51;
  Matches   31;  Conservative   22;  Mismatches   46;  Indels   18;  Gaps    5;

Qy       3 QKQEEENPAEETGEEKQDTQEKEG------ILPERAEEAKLKA----KYPSLGQKPGGSD 52
           ::: ||:|||   |:  | :| ||        |:  |  :  |    : || |::  ||:
Db     383 EEENEESPAENPQEQPSDGKENEGSEDQGSTPPDEEENEESPAEDPQEQPSDGKENKGSE 442

Qy      53 FLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSS 109
                :|    |  :   : |::  :| || | : |  : |   || |  : : |
Db     443 ------NQGSTPPDEEENEESPAEVPQEQ-PSDGKE-NEGSEDQGSTPPDEEENEES 491


RESULT  36
H86806
hypothetical protein yoiA [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: H86806
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: H86806
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1090 <STO>
A;Cross-references: GB:AE005176; PID:g12724447; PIDN:AAK05554.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: yoiA


  Query Match             12.8%;  Score 77.5;  DB 2;  Length 1090;
  Best Local Similarity   29.6%;  Pred. No. 79;
  Matches   32;  Conservative   17;  Mismatches   42;  Indels   17;  Gaps    5;

Qy       2 SQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGS------DFLM 55
           :::|||:  : |  : : | ::|:  || |  | :|      ||:    |       |
Db      20 TKEQEEKVLSAEIQKSQVDLKDKKETLPVRQREVEL-----MLGRNSYSSGLRTYKQLLE 74

Qy      56 KRLQKGQKY-FDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD 102
           | ||  :||  | |:|  |  |: |  |      | ||::       ||
Db      75 KYLQHSEKYSVDLGEYKKAYLKL-NSNL----KDFNLLSQKIADKKQD 117


RESULT  37
C71607
hypothetical protein PFB0745w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: C71607
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743
A;Accession: C71607
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1351 <GAR>
A;Cross-references: GB:AE001415; GB:AE001362; NID:g3845264; PIDN:AAC71938.1; PID:g384
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0745w


  Query Match             12.8%;  Score 77.5;  DB 2;  Length 1351;
  Best Local Similarity   30.8%;  Pred. No. 1e+02;
  Matches   24;  Conservative   15;  Mismatches   24;  Indels   15;  Gaps    3;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQ 62
           :|:|||:  :|  |||:: ::||    |  |: | | |   : :|      : | :||
Db     889 EKEEEEDKEKEKEEEKEEEKDKE----EEDEKDKEKEKEEEIQKK------VKKEIQKKV 938

Qy      63 KYFDSGDYNMAKAKMKNK 80
           |       |  | | :|:
Db     939 K-----KENQKKVKKENQ 951


RESULT  38
E89066
protein H05O09.1 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: E89066
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A;Accession: E89066
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2109 <STO>
A;Cross-references: GB:chr_V; PIDN:AAC19224.1; PID:g3193196; GSPDB:GN00023; CESP:H05O
C;Genetics:
A;Gene: H05O09.1
A;Map position: 5


  Query Match             12.8%;  Score 77.5;  DB 2;  Length 2109;
  Best Local Similarity   26.6%;  Pred. No. 1.6e+02;
  Matches   29;  Conservative   15;  Mismatches   30;  Indels   35;  Gaps    4;

Qy       6 EEENPAEETGEEKQDTQE--KEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           ||| | ||| |    | |  ||  :|| ||                            ||
Db    1859 EEETPLEETNETVVQTNEDVKEAEVPENAE--------------------------AQK 1891

Qy      64 YFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD-LPQRKSSLV 111
             || |  :| ::: :  :     |: ::   : |:  | | ::| |:|
Db    1892 VVDSSDLQVAASEIAHLAI-----DEAVLETSNQPSQFDSLQEQKPSVV 1935


RESULT  39
T33247
hypothetical protein H05O09.1 - Caenorhabditis elegans (fragment)
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33247
R;Leonard, S.; Hinds, K.
submitted to the EMBL Data Library, May 1998
A;Description: The sequence of C. elegans cosmid H05O09.
A;Reference number: Z21308
A;Accession: T33247
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2109 <LEO>
A;Cross-references: EMBL:AF067951; PIDN:AAC19224.1; GSPDB:GN00023; CESP:H05O09.1
A;Experimental source: strain Bristol N2; clone H05O09
C;Genetics:
A;Gene: CESP:H05O09.1
```

```
A;Map position: 5
A;Note: intron positions not resolved (incomplete sequence)


  Query Match             12.8%;  Score 77.5;  DB 2;  Length 2109;
  Best Local Similarity   26.6%;  Pred. No. 1.6e+02;
  Matches   29;  Conservative   15;  Mismatches   30;  Indels   35;  Gaps    4;

Qy       6 EEENPAEETGEEKQDTQE--KEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           ||| | ||| |    | |  ||  :|| ||                            ||
Db    1859 EEETPLEETNETVVQTNEDVKEAEVPENAE--------------------------AQK 1891

Qy      64 YFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD-LPQRKSSLV 111
             || |  :| ::: :  :     |: ::   : |:  | | ::| |:|
Db    1892 VVDSSDLQVAASEIAHLAI-----DEAVLETSNQPSQFDSLQEQKPSVV 1935


RESULT  40
AC1261
hypothetical protein lmo1491 [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AC1261
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AC1261
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-366 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC99569.1; PID:g16410920; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo1491


  Query Match             12.7%;  Score 77;  DB 2;  Length 366;
  Best Local Similarity   25.8%;  Pred. No. 27;
  Matches   24;  Conservative   18;  Mismatches   33;  Indels   18;  Gaps    3;

Qy      24 KEGILPERAEEAKL------KAKYPSL---------GQKPGGSDFLMKR---LQKGQKYF 65
           || :||:  :  ||      :|| |          :|  | | |:::   |: ||  :
Db     105 KEDVLPKSLKRDKLTRWMRTRAKEQGLAATDVVLVSAEKGHGFDTLLEKIDELRNGQDVY 164

Qy      66 DSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIP 98
             |  |: |: : |: :  |  : |::|    |
Db     165 VVGCTNVGKSTLINRIIKQASGENNVITTSQFP 197


Search completed: September  7, 2002, 10:44:25
Job time: 161 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:44:44 ; Search time 15.81 Seconds
(without alignments)
286.539 Million cell updates/sec

Title: US-09-824-178-3
Perfect score: 605
Sequence: 1 MSQKQEEENPAEETGEEKQD.....PTPQDLPQRKSSLVTSKLAG 117

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	427.5	70.7	111	1	AP19_HUMAN	P56211 homo sapien
2	427.5	70.7	111	1	AP19_MOUSE	P56212 mus musculu
3	425.5	70.3	111	1	AP19_BOVIN	Q28055 bos taurus
4	90.5	15.0	168	1	YNP7_YEAST	P53897 saccharomyc
5	83.5	13.8	407	1	IE68_HSVSA	Q01042 herpesvirus
6	82	13.6	644	1	NFM_RABIT	P54938 oryctolagus
7	81.5	13.5	583	1	SHC_HUMAN	P29353 homo sapien
8	81	13.4	155	1	H2B4_VOLCA	P16868 volvox cart
9	81	13.4	325	1	ZN36_HUMAN	P17029 homo sapien
10	81	13.4	1109	1	TCF8_RAT	Q62947 rattus norv
11	79	13.1	857	1	NFM_CHICK	P16053 gallus gall
12	79	13.1	1099	1	NKX1_HUMAN	O60721 homo sapien
13	79	13.1	4967	1	RYR2_HUMAN	Q92736 homo sapien
14	78.5	13.0	260	1	DH10_ARATH	P42759 arabidopsis
15	78.5	13.0	568	1	RGS3_MOUSE	Q9dc04 mus musculu
16	78.5	13.0	1359	1	ATRX_CAEEL	Q9u7e0 caenorhabdi
17	77	12.7	1220	1	IF2P_HUMAN	O60841 homo sapien
18	76.5	12.6	239	1	CALD_MELGA	P13505 meleagris g
19	76.5	12.6	771	1	CALD_CHICK	P12957 gallus gall
20	76.5	12.6	2004	1	MOZ_HUMAN	Q92794 homo sapien
21	75.5	12.5	300	1	TRT2_RABIT	P09741 oryctolagus
22	75.5	12.5	411	1	FKB3_YEAST	P38911 saccharomyc
23	75.5	12.5	790	1	SMY2_YEAST	P32909 saccharomyc
24	75.5	12.5	845	1	NFM_RAT	P12839 rattus norv
25	75.5	12.5	846	1	NRT_DROME	P23654 drosophila
26	75	12.4	357	1	FKB4_DROME	P54397 drosophila
27	75	12.4	877	1	INCE_CHICK	P53352 gallus gall
28	75	12.4	3924	1	ANK2_HUMAN	Q01484 homo sapien
29	74.5	12.3	243	1	NUKS_RAT	Q9epj0 rattus norv
30	74.5	12.3	851	1	DD21_MOUSE	Q9jik5 mus musculu
31	74.5	12.3	1739	1	CHD2_HUMAN	O14647 homo sapien
32	74.5	12.3	4128	1	PRKD_HUMAN	P78527 homo sapien
33	73.5	12.1	1054	1	RDGB_DROME	P43125 drosophila
34	73	12.1	467	1	HEX_ADE01	Q04965 human adeno
35	73	12.1	629	1	HAP1_RAT	P54256 rattus norv
36	73	12.1	705	1	TRDN_RABIT	Q28820 oryctolagus
37	73	12.1	807	1	YMT9_YEAST	Q04660 saccharomyc
38	73	12.1	1616	1	DNM1_HUMAN	P26358 homo sapien
39	73	12.1	2220	1	CABI_HUMAN	Q9y6j0 homo sapien
40	72.5	12.0	786	1	PRGR_CHICK	P07812 gallus gall
41	72.5	12.0	1411	1	TCOF_HUMAN	Q13428 homo sapien
42	72.5	12.0	1716	1	RPA1_RAT	O54889 rattus norv
43	72	11.9	219	1	BASP_RAT	Q05175 rattus norv
44	72	11.9	367	1	LHX4_HUMAN	Q969g2 homo sapien
45	72	11.9	573	1	PT1_BORBU	O51508 borrelia bu

ALIGNMENTS

RESULT 1
AP19_HUMAN
ID AP19_HUMAN STANDARD; PRT; 111 AA.
AC P56211; Q9UD70;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE cAMP-regulated phosphoprotein 19 (ARPP-19) [Contains: cAMP-regulated
DE phosphoprotein 16 (ARPP-16)].
GN ARPP19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98318667; PubMed=9653196;
RA Heron L., Virsolvy A., Peyrollier K., Gribble F.M., Ashcroft F.M.,
RA Lecam A., Bataille D.;
RT "Human alpha endosulfine, a possible regulator of sulphonylurea-
RT sensitive KATP channel. Molecular cloning, expression and biological
RT properties.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8387-8391(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20219739; PubMed=10754390;
RA Chin L.S., Singh S.K., Wang Q., Murray S.F.;
RT "Identification of okadaic-acid-induced genes by mRNA differential
RT display in glioma cells.";
RL J. Biomed. Sci. 7:152-159(2000).
RN [3]
RP SEQUENCE OF 16-111 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94165795; PubMed=8120638;
RA Brene S., Lindefors N., Ehrlich M., Taubes T., Horiuchi A., Kopp J.,
RA Hall H., Sedvall G., Greengard P., Persson H.;
RT "Expression of mRNAs encoding ARPP-16/19, ARPP-21, and DARPP-32 in
RT human brain tissue.";
RL J. Neurosci. 14:985-998(1994).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS ARE PROBABLY PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -!- PTM: SUBSTRATE FOR CAMP-DEPENDENT KINASE (BY SIMILARITY).
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; AJ223091; CAA11115.1; -.
DR EMBL; AF084555; AAD52044.1; -.

```
DR   MIM; 605487; -.
KW   Phosphorylation; Alternative splicing; Acetylation.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    111       CAMP-REGULATED PHOSPHOPROTEIN 19.
FT   CHAIN        16    111       CAMP-REGULATED PHOSPHOPROTEIN 16.
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
FT   MOD_RES      16     16       ACETYLATION (BY SIMILARITY).
SQ   SEQUENCE   111 AA;  12192 MW;  357D3F039EDCABC1 CRC64;


  Query Match           70.7%;  Score 427.5;  DB 1;  Length 111;
  Best Local Similarity 76.4%;  Pred. No. 5.2e-30;
  Matches   84;  Conservative    9;  Mismatches   16;  Indels    1;  Gaps    1;

Qy      8 ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDS 67
          | |   : ||::: ::|    ||:||||||||:|| |||||||||| ||||||||||
Db      3 EVPEAASAEEQKEMEDKV-TSPEKAEEAKLKARYPHLGQKPGGSDFLRKRLQKGQKYFDS 61

Qy     68 GDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
          |||||||||||||||:| |||  |||||||||||||||| ||| |||||
Db     62 GDYNMAKAKMKNKQLPTAAPDKTEVTGDHIPTPQDLPQRKPSLVASKLAG 111


RESULT   2
AP19_MOUSE
ID   AP19_MOUSE     STANDARD;      PRT;   111 AA.
AC   P56212;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   cAMP-regulated phosphoprotein 19 (ARPP-19).
GN   ARPP19.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,
RA   Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,
RA   Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B.,
RA   Theising B., Wylie T., Lennon G., Soares B., Wilson R., Waterston R.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- PTM: SUBSTRATE FOR CAMP-DEPENDENT KINASE (BY SIMILARITY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AA260149; -; NOT_ANNOTATED_CDS.
DR   EMBL; W87203; -; NOT_ANNOTATED_CDS.
DR   EMBL; AA060186; -; NOT_ANNOTATED_CDS.
DR   EMBL; W70699; -; NOT_ANNOTATED_CDS.
DR   EMBL; AA041700; -; NOT_ANNOTATED_CDS.
DR   EMBL; AA500116; -; NOT_ANNOTATED_CDS.
DR   EMBL; W66839; -; NOT_ANNOTATED_CDS.
DR   MGD; MGI:1891691; Arpp19.
KW   Phosphorylation; Acetylation.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
SQ   SEQUENCE   111 AA;  12162 MW;  356089D8EEDCABC1 CRC64;


  Query Match           70.7%;  Score 427.5;  DB 1;  Length 111;
  Best Local Similarity 76.4%;  Pred. No. 5.2e-30;
  Matches   84;  Conservative    9;  Mismatches   16;  Indels    1;  Gaps    1;

Qy      8 ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDS 67
          | |   : ||::: ::|    ||:||||||||:|| |||||||||| ||||||||||
Db      3 EVPEAASAEEQKEMEDKV-TSPEKAEEAKLKARYPHLGQKPGGSDFLRKRLQKGQKYFDS 61

Qy     68 GDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
          |||||||||||||||:| |||  |||||||||||||||| ||| |||||
Db     62 GDYNMAKAKMKNKQLPAAAPDKTEVTGDHIPTPQDLPQRKPSLVASKLAG 111


RESULT   3
AP19_BOVIN
ID   AP19_BOVIN     STANDARD;      PRT;   111 AA.
AC   Q28055; Q28054;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   cAMP-regulated phosphoprotein 19 (ARPP-19) [Contains: cAMP-regulated
DE   phosphoprotein 16 (ARPP-16)].
GN   ARPP19.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=90264450; PubMed=2160982;
RA   Horiuchi A., Williams K.R., Kurihara T., Nairn A.C., Greengard P.;
RT   "Purification and cDNA cloning of ARPP-16, a cAMP-regulated
RT   phosphoprotein enriched in basal ganglia, and of a related
RT   phosphoprotein, ARPP-19.";
RL   J. Biol. Chem. 265:9476-9484(1990).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS ARE PROBABLY PRODUCED BY
CC       ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: ARPP-19 FOUND IN ALL BRAIN REGIONS AND ALSO
CC       PRESENT IN NON-NEURONAL TISSUES. ARPP-16 ENRICHED IN THE CAUDATE
CC       NUCLEUS, FOUND IN LOW LEVELS IN CEREBRAL CORTEX.
CC   -!- PTM: SUBSTRATE FOR CAMP-DEPENDENT KINASE.
CC   -!- MASS SPECTROMETRY: MW=10708; MW_ERR=1; METHOD=MALDI;
CC       RANGE=16-111.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33618; AAA30386.1; -.
DR   EMBL; M33617; AAA30385.1; -.
KW   Phosphorylation; Alternative splicing; Acetylation.
FT   INIT_MET      0      0
FT   CHAIN         1    111       CAMP-REGULATED PHOSPHOPROTEIN 19.
FT   CHAIN        16    111       CAMP-REGULATED PHOSPHOPROTEIN 16.
FT   MOD_RES       1      1       ACETYLATION.
FT   MOD_RES      16     16       ACETYLATION.
SQ   SEQUENCE   111 AA;  12222 MW;  43A64F039EDCABD5 CRC64;


  Query Match           70.3%;  Score 425.5;  DB 1;  Length 111;
  Best Local Similarity 76.4%;  Pred. No. 7.6e-30;
  Matches   84;  Conservative    9;  Mismatches   16;  Indels    1;  Gaps    1;

Qy      8 ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDS 67
          | |   : ||::: ::|    ||:||||||||:|| |||||||||| ||||||||||
Db      3 EVPEAASAEEQKEMEDKV-TSPEKAEEAKLKARYPHLGQKPGGSDFLRKRLQKGQKYFDS 61

Qy     68 GDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
          |||||||||||||||:| |||  |||||||||||||||| ||| |||||
Db     62 GDYNMAKAKMKNKQLPTATPDKTEVTGDHIPTPQDLPQRKPSLVASKLAG 111
```

```
RESULT   4
YNP7_YEAST
ID   YNP7_YEAST     STANDARD;      PRT;   168 AA.
AC   P53897;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Hypothetical 18.1 kDa protein in YGP1-YCK2 intergenic region.
GN   YNL157W OR N1743.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE=96287653; PubMed=8686380;
RA   Nasr F., Becam A.-M., Herbert C.J.;
RT   "The sequence of 36.8 kb from the left arm of chromosome XIV reveals
RT   24 complete open reading frames: 18 correspond to new genes, one of
RT   which encodes a protein similar to the human myotonic dystrophy
RT   kinase.";
RL   Yeast 12:169-175(1996).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X92517; CAA63282.1; -.
DR   EMBL; Z71433; CAA96044.1; -.
DR   SGD; S0005101; YNL157W.
KW   Hypothetical protein.
FT   DOMAIN      155    160       POLY-SER.
SQ   SEQUENCE   168 AA;  18051 MW;  9534A95F1BD2CFFB CRC64;


  Query Match             15.0%;  Score 90.5;  DB 1;  Length 168;
  Best Local Similarity   31.9%;  Pred. No. 0.32;
  Matches   30;  Conservative   11;  Mismatches   30;  Indels   23;  Gaps    5;

Qy      34 EAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKN--------KQLPSA 85
           | ||   |   |: |   |     :|| :|||||||| : || ::|          ||
Db      35 EMKL---YKMYGKLPSKKDIFKHTMQK-RKYFDSGDYALQKAGIQNNDPINYGKNNLPLT 90

Qy      86 GPDKNLVTGDHIPTPQDLPQRKSSLV--TSKLAG 117
             | |           :|: :|: |    |:  ||
Db      91 NPSK---------LREDIIKRRISTCPSTASTAG 115


RESULT   5
IE68_HSVSA
ID   IE68_HSVSA     STANDARD;      PRT;   407 AA.
AC   Q01042;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   Immediate-early protein.
GN   73 OR ECLF1.
OS   Herpesvirus saimiri (strain 11).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=10383;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92333688; PubMed=1321287;
RA   Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA   Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA   Honess R.W.;
RT   "Primary structure of the herpesvirus saimiri genome.";
RL   J. Virol. 66:5047-5058(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92230228; PubMed=1314457;
RA   Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;
RT   "Analysis of nucleotide sequence of the rightmost 43 kbp of
RT   herpesvirus saimiri (HVS) L-DNA: general conservation of genetic
RT   organization between HVS and Epstein-Barr virus.";
RL   Virology 188:296-310(1992).
CC   -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-1 AND
CC       HSV-2 IE-68 (US1), EHV-1 65, EHV-4 (ORF4), PRV RSP40, AND VZV 63.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64346; CAA45696.1; -.
DR   EMBL; M86409; AAA46149.1; -.
DR   EMBL; S76368; AAB21116.1; -.
DR   PIR; G36813; EDBEQ3.
DR   PIR; S20244; S20244.
DR   HSSP; P53041; 1A17.
KW   Early protein.
FT   DOMAIN      60    241       GLU-RICH (ACIDIC).
SQ   SEQUENCE   407 AA;  46617 MW;  FFD399CA82CE136C CRC64;


  Query Match             13.8%;  Score 83.5;  DB 1;  Length 407;
  Best Local Similarity   30.6%;  Pred. No. 3.2;
  Matches   33;  Conservative   13;  Mismatches   35;  Indels   27;  Gaps    5;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAK----------YPSLGQK---- 47
           : :||   |||  || :: :| |    | ||||: : :           |  :|||
Db     203 EAEEEAEEAEEEAEEAEEAEEAEE-AEEEAEEEAEEEEEEAGPSTPRLPHYKVVGQKPSTQ 261

Qy      48 PGGSDFL--------MKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGP 87
           |||   |          || ||::   | |    | : :||    | |
Db     262 PGGVPKLCLKMQPQHRSRLPKGKQ---SHDKVPKKYQARNKFFSQAAP 306


RESULT   6
NFM_RABIT
ID   NFM_RABIT      STANDARD;      PRT;   644 AA.
AC   P54938;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neurofilament triplet M protein (160 kDa neurofilament protein)
DE   (Neurofilament medium polypeptide) (NF-M) (Fragment).
GN   NEF3 OR NEFM OR NFM.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=97055255; PubMed=8899542;
RA   Vitadello M., Vettore S., Lamar E., Chien K.R., Gorza L.;
RT   "Neurofilament M mRNA is expressed in conduction system myocytes of
RT   the developing and adult rabbit heart.";
RL   J. Mol. Cell. Cardiol. 28:1833-1844(1996).
CC   -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC       AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC   -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
CC       PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC       THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC       INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC       OF AXONAL CALIBER.
```

```
CC   -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC       OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC       LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC       COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z47378; CAA87454.1; -.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW   Glycoprotein.
FT   NON_TER       1      1
FT   DOMAIN       <1    197       ROD.
FT   DOMAIN      198    643       TAIL.
FT   DOMAIN       <1     33       COIL 1B.
FT   DOMAIN       34     50       LINKER 12.
FT   DOMAIN       51     72       COIL 2A.
FT   DOMAIN       73     76       LINKER 2.
FT   DOMAIN       77    197       COIL 2B.
FT   CARBOHYD    217    217       O-LINKED (GLCNAC) (BY SIMILARITY).
SQ   SEQUENCE   644 AA;  72450 MW;  030FDAA622889678 CRC64;


  Query Match             13.6%;  Score 82;  DB 1;  Length 644;
  Best Local Similarity   28.9%;  Pred. No. 7.1;
  Matches   22;  Conservative   17;  Mismatches   27;  Indels   10;  Gaps    2;

Qy       5 QEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKY 64
           :|||   || |:|::: :| ||:  ::|||          | :  ||       ::|:
Db     308 KEEEGEKEEEGQEEEEEEEDEGVKSDQAEEG---------GSEKEGSSKNEGEQEEGETE 358

Qy      65 FDSGDYNMAKAKMKNK 80
             : |:   |:|| : |
Db     359 AE-GEVEEAEAKEEKK 373


RESULT   7
SHC_HUMAN
ID   SHC_HUMAN      STANDARD;      PRT;   583 AA.
AC   P29353; O15290;
DT   01-DEC-1992 (Rel. 24, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   SHC transforming protein.
GN   SHC1 OR SHC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (P46SHC AND P52SHC ISOFORMS).
RX   MEDLINE=92323554; PubMed=1623525;
RA   Pelicci G., Lanfrancone L., Grignani F., McGlade J., Cavallo F.,
RA   Forni G., Nicoletti I., Grignani F., Pawson T., Pelicci P.G.;
RT   "A novel transforming protein (SHC) with an SH2 domain is implicated
RT   in mitogenic signal transduction.";
RL   Cell 70:93-104(1992).
RN   [2]
RP   SEQUENCE FROM N.A. (P66SHC ISOFORM).
RX   MEDLINE=97201514; PubMed=9049300;
RA   Migliaccio E., Mele S., Salcini A.E., Pelicci G., Lai K.M.,
RA   Superti-Furga G., Pawson T., Di Fiore P.P., Lanfrancone L.,
RA   Pelicci P.G.;
RT   "Opposite effects of the p52shc/p46shc and p66shc splicing isoforms on
RT   the EGF receptor-MAP kinase-fos signalling pathway.";
RL   EMBO J. 16:706-716(1997).
RN   [3]
RP   SEQUENCE FROM N.A. (P66SHC ISOFORM).
RA   Harun R., Smith K.K., Leek J.P., Markham A.F., Norris A.,
RA   Morrison J.F.J.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 482-583.
RX   MEDLINE=96080377; PubMed=7473762;
RA   Mikol V., Baumann G., Zurini M.G.M., Hommel U.;
RT   "Crystal structure of the SH2 domain from the adaptor protein SHC: a
RT   model for peptide binding based on X-ray and NMR data.";
RL   J. Mol. Biol. 254:86-95(1995).
RN   [5]
RP   STRUCTURE BY NMR OF 127-317.
RX   MEDLINE=96097066; PubMed=8524391;
RA   Zhou M.M., Ravichandran K.S., Olejniczak E.F., Petros A.M.,
RA   Meadows R.P., Sattler M., Harlan J.E., Wade W.S., Burakoff S.J.,
RA   Fesik S.W.;
RT   "Structure and ligand recognition of the phosphotyrosine binding
RT   domain of Shc.";
RL   Nature 378:584-592(1995).
CC   -!- FUNCTION: MAY COUPLE ACTIVATED GROWTH FACTOR RECEPTORS TO A
CC       SIGNALING PATHWAY THAT REGULATES THE PROLIFERATION OF MAMMALIAN
CC       CELLS. SHC MIGHT PARTICIPATE IN THE TRANSFORMING ACTIVITY OF
CC       ONCOGENIC TYROSINE KINASES.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- ALTERNATIVE PRODUCTS: 3 isoforms; p66shc (shown here), p52shc
CC       and p46shc; are produced by alternative initiation.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED IN MAMMALIAN TISSUES.
CC   -!- DOMAIN: THE PID DOMAIN SPECIFICALLY BINDS TO THE ASN-PRO-XAA-
CC       TYR(P) MOTIF FOUND IN MANY TYROSINE-PHOSPHORYLATED PROTEINS
CC       INCLUDING GROWTH FACTOR RECEPTORS.
CC   -!- PTM: PHOSPHORYLATED BY ACTIVATED EPIDERMAL GROWTH FACTOR
CC       RECEPTOR.
CC   -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PID DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X68148; CAA48251.1; -.
DR   EMBL; U73377; AAB49972.1; -.
DR   EMBL; Y09847; CAA70977.1; -.
DR   PIR; S25776; S25776.
DR   PDB; 1MIL; 08-NOV-96.
DR   PDB; 1SHC; 15-MAY-97.
DR   MIM; 600560; -.
DR   InterPro; IPR000050; PID_domain.
DR   InterPro; IPR000980; SH2.
DR   Pfam; PF00640; PID; 1.
DR   Pfam; PF00017; SH2; 1.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   PRINTS; PR00629; SHCPIDOMAIN.
DR   SMART; SM00462; PTB; 1.
DR   SMART; SM00252; SH2; 1.
DR   PROSITE; PS01179; PID; 1.
DR   PROSITE; PS50001; SH2; 1.
KW   SH2 domain; Phosphorylation; Growth regulation; Proto-oncogene;
KW   Alternative initiation; 3D-structure.
FT   CHAIN         1    583       SHC TRANSFORMING PROTEIN, P66SHC
FT                                ISOFORM.
FT   CHAIN       111    583       SHC TRANSFORMING PROTEIN, P52SHC
FT                                ISOFORM.
FT   CHAIN       156    583       SHC TRANSFORMING PROTEIN, P46SHC
FT                                ISOFORM.
FT   INIT_MET    111    111       FOR P52SHC ISOFORM.
```

```
FT   INIT_MET    156    156       FOR P46SHC ISOFORM.
FT   DOMAIN      156    339       PID.
FT   DOMAIN      488    579       SH2.
FT   CONFLICT      2      2       D -> N (IN REF. 3).
FT   CONFLICT     21     21       L -> M (IN REF. 3).
FT   CONFLICT     38     38       S -> P (IN REF. 3).
FT   CONFLICT     95     95       D -> V (IN REF. 3).
FT   CONFLICT    101    101       E -> D (IN REF. 3).
SQ   SEQUENCE   583 AA;  62852 MW;  8B07E8806711C933 CRC64;
```

```
  Query Match             13.5%;  Score 81.5;  DB 1;  Length 583;
  Best Local Similarity   26.2%;  Pred. No. 7;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   33;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|       | : :||  |  |      |: ||       :||
Db     339 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 398

Qy      47 KPGGSDFLMKRLQ-----KGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDHI 97
             ||    : |::       |:: ||   |    |: ||:   : :  |||    : |
Db     399 PVGGDPEVRKQMPPPPPCPGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS-- 453

Qy      98 PTPQDL 103
             |:||
Db     454 -APRDL 458
```

```
RESULT   8
H2B4_VOLCA
ID   H2B4_VOLCA     STANDARD;      PRT;   155 AA.
AC   P16868;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Histone H2B-IV.
OS   Volvox carteri.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Volvocaceae; Volvox.
OX   NCBI_TaxID=3067;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F. NAGARIENSIS / HK10;
RX   MEDLINE=91033024; PubMed=2227431;
RA   Mueller K., Lindauer A., Bruederlein M., Schmitt R.;
RT   "Organization and transcription of Volvox histone-encoding genes:
RT   similarities between algal and animal genes.";
RL   Gene 93:167-175(1990).
CC   -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC       H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H2B FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M31922; AAA34250.1; -.
DR   PIR; JQ0797; JQ0797.
DR   InterPro; IPR000558; Histone_H2B.
DR   InterPro; IPR000166; Histone_core.
DR   Pfam; PF00125; histone; 1.
DR   PRINTS; PR00621; HISTONEH2B.
DR   ProDom; PD000497; Histone_H2B; 1.
DR   SMART; SM00427; H2B; 1.
DR   PROSITE; PS00357; HISTONE_H2B; 1.
KW   Nuclear protein; Chromosomal protein; Nucleosome core; DNA-binding;
KW   Multigene family.
SQ   SEQUENCE   155 AA;  16841 MW;  56AC3D9C7A6F5940 CRC64;
```

```
  Query Match             13.4%;  Score 81;  DB 1;  Length 155;
  Best Local Similarity   29.2%;  Pred. No. 1.9;
  Matches   26;  Conservative   13;  Mismatches   42;  Indels    8;  Gaps    2;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEE---AKLKAKYPSLGQKPGGSDFLMKR 57
           |: | :|| || |  | | : : |    |:: |:   ||  || |: |    |      :
Db       1 MAPKTKEEKPASEAVEPKAEAKPKAEKAPKKKEKKAPAKKSAKEPAAGDAAEGD-----K 55

Qy      58 LQKGQKYFDSGDYNMAKAKMKNKQLPSAG 86
             :|  |   |  | :   |:  :  |  |
Db      56 KKKKAKVAKSETYKLYIYKVLKQVHPDTG 84
```

```
RESULT   9
ZN36_HUMAN
ID   ZN36_HUMAN     STANDARD;      PRT;   325 AA.
AC   P17029; P52745;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Zinc finger protein 36 (Zinc finger protein KOX18) (Fragment).
GN   ZNF36 OR KOX18 OR ZNF139.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Insulinoma;
RX   MEDLINE=96044430; PubMed=7557990;
RA   Tommerup N., Vissing H.;
RT   "Isolation and fine mapping of 16 novel human zinc finger-encoding
RT   cDNAs identify putative candidate genes for developmental and
RT   malignant disorders.";
RL   Genomics 27:259-264(1995).
RN   [2]
RP   SEQUENCE OF 167-222 FROM N.A.
RC   TISSUE=Lymphoid;
RX   MEDLINE=91145339; PubMed=2288909;
RA   Thiesen H.-J.;
RT   "Multiple genes encoding zinc finger domains are expressed in human T
RT   cells.";
RL   New Biol. 2:363-374(1990).
CC   -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U09848; AAC50264.1; -.
DR   EMBL; X52349; CAA36575.1; -.
DR   PIR; S10406; S10406.
DR   HSSP; P25490; 1ZNM.
DR   MIM; 601260; -.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00096; zf-C2H2; 6.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00355; ZnF_C2H2; 6.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 6.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 6.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Repeat.
FT   NON_TER       1      1
FT   DOMAIN      139    301       ZINC FINGERS.
```

```
FT   ZN_FING      139    161       C2H2-TYPE.
FT   ZN_FING      167    189       C2H2-TYPE.
FT   ZN_FING      195    217       C2H2-TYPE.
FT   ZN_FING      223    245       C2H2-TYPE.
FT   ZN_FING      251    273       C2H2-TYPE.
FT   ZN_FING      279    301       C2H2-TYPE.
SQ   SEQUENCE   325 AA;  36640 MW;  2659C09BD7B05708 CRC64;


  Query Match             13.4%;  Score 81;  DB 1;  Length 325;
  Best Local Similarity   21.3%;  Pred. No. 4.1;
  Matches   32;  Conservative   25;  Mismatches   47;  Indels   46;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           :||    |   | | ::  |:     | :|::  : :     ||    :|  || |:|:
Db      18 RQENYGSAFPQGGENRNENEESTSKAETSEDSASRGETTGRSQK----EFGEKRDQEGK- 72

Qy      64 YFDSGDYNMAKAKMKNKQL-----PSAGPDKNLVTGDHIP-----------------TP 100
             :|:      : | ::      |: | ||  :||:  |                   ||
Db      73 ---TGERQQKNPEEKTRKEKRDSGPAIGKDKKTITGERGPREKGKGLGRSFSLSSNFTTP 129

Qy     101 QDLP--------------QRKSSLVTSKL 115
           :::|               | |||:  |:
Db     130 EEVPTGTKSHRCDECGKCFTRSSSLIRHKI 159


RESULT  10
TCF8_RAT
ID   TCF8_RAT       STANDARD;      PRT;  1109 AA.
AC   Q62947; Q62948;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Transcription factor 8 (Zinc finger homeodomain enhancer-binding
DE   protein) (Zfhep).
GN   TCF8.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 6-1109 FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=96365389; PubMed=8769566;
RA   Cabanillas A.M., Darling D.S.;
RT   "Alternative splicing gives rise to two isoforms of Zfhep, a zinc
RT   finger/homeodomain protein that binds T3-response elements.";
RL   DNA Cell Biol. 15:643-651(1996).
CC   -!- FUNCTION: ACTS AS A TRANSCRIPTIONAL REPRESSOR. BINDS TO E-BOX
CC       SEQUENCES IN THE IMMUNOGLOBULIN HEAVY CHAIN ENHANCER AS WELL AS IN
CC       THE REGULATORY REGIONS OF MANY OTHER TISSUE-SPECIFIC GENES (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1/ZFHEP-1 (SHOWN HERE) AND
CC       2/ZFHEP-2; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO DELTA-EF1/ZFH-1 FAMILY OF TWO-HANDED ZINC
CC       FINGER/HOMEODOMAIN PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U51583; AAB17130.1; -.
DR   EMBL; U51584; AAB17131.1; -.
DR   HSSP; P08046; 1A1G.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF00096; zf-C2H2; 7.
DR   SMART; SM00389; HOX; 1.
DR   SMART; SM00355; ZnF_C2H2; 7.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 6.
KW   Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
KW   Homeobox; Repressor; Activator; Metal-binding; Repeat;
KW   Alternative splicing.
FT   ZN_FING      150    173       C2H2-TYPE.
FT   ZN_FING      180    202       C2H2-TYPE.
FT   ZN_FING      220    242       C2H2-TYPE.
FT   ZN_FING      248    272       C2H2-TYPE (ATYPICAL).
FT   DNA_BIND     559    618       HOMEOBOX-LIKE.
FT   ZN_FING      881    903       C2H2-TYPE.
FT   ZN_FING      909    931       C2H2-TYPE.
FT   ZN_FING      937    958       C2H2-TYPE (ATYPICAL).
FT   DOMAIN       968   1109       GLU-RICH (ACIDIC).
FT   VARSPLIC       1    198       MISSING (IN ISOFORM 2).
SQ   SEQUENCE  1109 AA;  121626 MW;  BEFE291C8795DDA6 CRC64;


  Query Match             13.4%;  Score 81;  DB 1;  Length 1109;
  Best Local Similarity   33.3%;  Pred. No. 15;
  Matches   26;  Conservative   11;  Mismatches   21;  Indels   20;  Gaps    3;

Qy       1 MSQKQEE---ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLG------------ 45
           | : ||:   ||| ||  ||::: :|:|    | ||||: :|     |
Db    1028 MEELQEDKECENPQEEEEEEEEEEEEEEEEEEEEEAEEAEHEAAAAKTGGAVEEEAAQQAG 1087

Qy      46 ---QKPGGSDFLMKRLQK 60
             .  ||  ||:   ||| :
Db    1088 SFQQKASGSE--SKRLSE 1103


RESULT  11
NFM_CHICK
ID   NFM_CHICK       STANDARD;      PRT;   857 AA.
AC   P16053;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neurofilament triplet M protein (160 kDa neurofilament protein)
DE   (Neurofilament medium polypeptide) (NF-M).
GN   NEFM.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90174973; PubMed=2106668;
RA   Zopf D., Dineva B., Betz H., Gundelfinger E.D.;
RT   "Isolation of the chicken middle-molecular weight neurofilament
RT   (NF-M) gene and characterization of its promoter.";
RL   Nucleic Acids Res. 18:521-529(1990).
RN   [2]
RP   SEQUENCE OF 259-857 FROM N.A.
RX   MEDLINE=88112814; PubMed=3123320;
RA   Zopf D., Hermans-Borgmeyer I., Gundelfinger E.D., Betz H.;
RT   "Identification of gene products expressed in the developing chick
RT   visual system: characterization of a middle-molecular-weight
RT   neurofilament cDNA.";
RL   Genes Dev. 1:699-708(1987).
CC   -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC       AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC   -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
CC       PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC       THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC       INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC       OF AXONAL CALIBER.
CC   -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC       OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC       LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC       COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
```

```
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X17102; CAA34958.1; -.
DR   EMBL; X05558; CAA29073.1; -.
DR   PIR; A27040; A27040.
DR   PIR; S08061; S08061.
DR   PIR; S15762; S15762.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW   Glycoprotein.
FT   INIT_MET      0      0
FT   DOMAIN        1     98       HEAD.
FT   DOMAIN       99    406       ROD.
FT   DOMAIN      407    857       TAIL.
FT   DOMAIN       99    130       COIL 1A.
FT   DOMAIN      131    143       LINKER 1.
FT   DOMAIN      144    242       COIL 1B.
FT   DOMAIN      243    259       LINKER 12.
FT   DOMAIN      260    281       COIL 2A.
FT   DOMAIN      282    285       LINKER 2.
FT   DOMAIN      286    406       COIL 2B.
FT   CARBOHYD     46     46       O-LINKED (GLCNAC) (BY SIMILARITY).
FT   CARBOHYD    426    426       O-LINKED (GLCNAC) (BY SIMILARITY).
FT   CONFLICT    546    546       G -> R (IN REF. 2).
SQ   SEQUENCE   857 AA;  95704 MW;  4E2E0FC6AC64778B CRC64;


  Query Match             13.1%;  Score 79;  DB 1;  Length 857;
  Best Local Similarity   25.2%;  Pred. No. 17;
  Matches   36;  Conservative   18;  Mismatches   59;  Indels   30;  Gaps    5;

Qy       3 QKQEEE---------NPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQK------ 47
           :|:|||         : ||| | :|:: :|||       |||: | |    | |
Db     513 EKEEEEAEEEEAAKSDAAEEGGSKKEEIEEKEEGEEAEEEEAEAKGKAEEAGAKVEKVKS 572

Qy      48 ---------PGGSDFL--MKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNL---VT 93
                  |  |      | :||              |:   |:  :| |:|     ||
Db     573 PPAKSPPKSPPKSPVTEQAKAVQKAAAEVGKDQKAEKAAEKAAKEEKAASPEKPATPKVT 632

Qy      94 GDHIP-TPQDLPQRKSSLVTSKL 115
             | ||:  |  : ::   |:
Db     633 SPEKPATPEKPPTPEKAITPEKV 655


RESULT  12
NKX1_HUMAN
ID   NKX1_HUMAN     STANDARD;      PRT;  1099 AA.
AC   O60721; O43485; O75184;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Sodium/potassium/calcium exchanger 1 precursor (Na(+)/K(+)/Ca(2+)-
DE   exchange protein 1) (Retinal rod Na-Ca+K exchanger).
GN   SLC24A1 OR NCKX1 OR KIAA0702.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Retina;
RX   MEDLINE=99072302; PubMed=9856482;
RA   Tucker J.E., Winkfein R.J., Murthy S.K., Friedman J.S., Walter M.A.,
RA   Demetrick D.J., Schnetkamp P.P.M.;
RT   "Chromosomal localization and genomic organization of the human
RT   retinal rod Na-Ca+K exchanger.";
RL   Hum. Genet. 103:411-414(1998).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Retina;
RX   MEDLINE=98138491; PubMed=9478004;
RA   Tucker J.E., Winkfein R.J., Cooper C.B., Schnetkamp P.P.M.;
RT   "cDNA cloning of the human retinal rod Na-Ca+K exchanger: comparison
RT   with a revised bovine sequence.";
RL   Invest. Ophthalmol. Vis. Sci. 39:435-440(1998).
RN   [3]
RP   SEQUENCE OF 1-999 FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811;
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
CC   -!- FUNCTION: Critical component of the visual transduction cascade,
CC       controlling the calcium concentration of outer segments during
CC       light and darkness. Light causes a rapid lowering of cytosolic
CC       free calcium in the outer segment of both retinal rod and cone
CC       photoreceptors and the light-induced lowering of calcium is caused
CC       by extrusion via this protein which plays a key role in the
CC       process of light adaptation. Transports one Ca(2+) and one K(+) in
CC       exchange for four Na(+).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
CC       produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Found only in the outer segments of retinal
CC       rod photoreceptors.
CC   -!- SIMILARITY: BELONGS TO THE SLC24A FAMILY OF TRANSPORTERS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF062921; AAC16732.1; -.
DR   EMBL; AF062922; AAC77912.1; -.
DR   EMBL; AF026132; AAB97832.1; -.
DR   EMBL; AB014602; BAA31677.1; ALT_SEQ.
DR   MIM; 603617; -.
DR   InterPro; IPR002613; Na_Ca_Ex.
DR   Pfam; PF01699; Na_Ca_Ex; 2.
KW   Vision; Transport; Symport; Transmembrane; Glycoprotein;
KW   Phosphorylation; Signal; Calcium transport; Repeat;
KW   Alternative splicing.
FT   SIGNAL        1     38       POTENTIAL.
FT   CHAIN        39   1099       SODIUM/POTASSIUM/CALCIUM EXCHANGER 1.
FT   DOMAIN       39    452       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    453    473       POTENTIAL.
FT   DOMAIN      474    497       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    498    518       POTENTIAL.
FT   DOMAIN      519    522       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    523    543       POTENTIAL.
FT   DOMAIN      544    563       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    564    584       POTENTIAL.
FT   DOMAIN      585    585       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    586    606       POTENTIAL.
FT   DOMAIN      607    907       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    908    928       POTENTIAL.
FT   DOMAIN      929    935       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    936    956       POTENTIAL.
FT   DOMAIN      957    971       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    972    992       POTENTIAL.
```

```
FT   DOMAIN      993   1010       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1011   1031       POTENTIAL.
FT   DOMAIN     1032   1039       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1040   1060       POTENTIAL.
FT   DOMAIN     1061   1068       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1069   1089       POTENTIAL.
FT   DOMAIN     1090   1099       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      494    534       ALPHA-1.
FT   REPEAT      979   1010       ALPHA-2.
FT   DOMAIN      860    890       POLY-GLU.
FT   MOD_RES     658    658       PHOSPHORYLATION (POTENTIAL).
FT   CARBOHYD    290    290       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    631    648       MISSING (IN ISOFORM 2).
FT   CONFLICT    516    516       V -> I (IN REF. 2).
SQ   SEQUENCE   1099 AA;  121374 MW;  3911856BB088B5FD CRC64;


  Query Match             13.1%;  Score 79;  DB 1;  Length 1099;
  Best Local Similarity   28.0%;  Pred. No. 22;
  Matches   21;  Conservative   14;  Mismatches   28;  Indels   12;  Gaps    1;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMK------ 56
           :::|||   ||  ||::: :|:|    |:  |  |   :|   ||     ||:
Db     863 EEEEEEQEEEEEEEEQEEEEEEEEEEEEKGNEEPLSLDWPETRQKQAIYLFLLPIVFPLW 922

Qy      57 ------RLQKGQKYF 65
                 | |: :|:|
Db     923 LTVPDVRRQESRKFF 937


RESULT  13
RYR2_HUMAN
ID   RYR2_HUMAN     STANDARD;      PRT;  4967 AA.
AC   Q92736; Q15411;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Ryanodine receptor 2 (Cardiac muscle-type ryanodine receptor) (RyR2)
DE   (RYR-2) (Cardiac muscle ryanodine receptor-calcium release channel)
DE   (hRYR-2).
GN   RYR2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Heart muscle;
RX   MEDLINE=96404895; PubMed=8809036;
RA   Tunwell R.E.A., Wickenden C., Bertrand B.M.A., Shevchenko V.I.,
RA   Walsh M.B., Allen P.D., Lai F.A.;
RT   "The human cardiac muscle ryanodine receptor-calcium release channel:
RT   identification, primary structure and topological analysis.";
RL   Biochem. J. 318:477-487(1996).
RN   [2]
RP   SEQUENCE OF 9-87 AND 533-681 FROM N.A., DEVELOPMENTAL STAGE, AND
RP   INDUCTION.
RC   TISSUE=Myometrium, and Heart muscle;
RX   MEDLINE=97220386; PubMed=9148749;
RA   Awad S.S., Lamb H.K., Morgan J.M., Dunlop W., Gillespie J.I.;
RT   "Differential expression of ryanodine receptor RyR2 mRNA in the non-
RT   pregnant and pregnant human myometrium.";
RL   Biochem. J. 322:777-783(1997).
RN   [3]
RP   SEQUENCE OF 4292-4479 FROM N.A., AND TISSUE SPECIFICITY.
RC   TISSUE=Cerebellum, and Hippocampus;
RX   MEDLINE=98268728; PubMed=9607712;
RA   Martin C., Chapman K.E., Seckl J.R., Ashley R.H.;
RT   "Partial cloning and differential expression of ryanodine
RT   receptor/calcium-release channel genes in human tissues including the
RT   hippocampus and cerebellum.";
RL   Neuroscience 85:205-216(1998).
RN   [4]
RP   VARIANTS VTSIP L-2246; S-2474; K-4104 AND C-4497.
RX   PubMed=11208676;
RA   Priori S.G., Napolitano C., Tiso N., Memmi M., Vignati G., Bloise R.,
RA   Sorrentino V.V., Danieli G.A.;
RT   "Mutations in the cardiac ryanodine receptor gene (hRyR2) underlie
RT   catecholaminergic polymorphic ventricular tachycardia.";
RL   Circulation 103:196-200(2001).
RN   [5]
RP   VARIANTS VPFT S-2328; R-4201 AND F-4653, AND VARIANT R-2958.
RX   MEDLINE=21112862; PubMed=11157710;
RA   Laitinen P.J., Brown K.M., Piippo K., Swan H., Devaney J.M.,
RA   Brahmbhatt B., Donarum E.A., Marino M., Tiso N., Viitasalo M.,
RA   Toivonen L., Stephan D.A., Kontula K.;
RT   "Mutations of the cardiac ryanodine receptor (RyR2) gene in familial
RT   polymorphic ventricular tachycardia.";
RL   Circulation 103:485-490(2001).
RN   [6]
RP   VARIANTS ARVD2 PRO-433 AND ILE-2386.
RX   MEDLINE=21096894; PubMed=11159936;
RA   Tiso N., Stephan D.A., Nava A., Bagattin A., Devaney J.M., Stanchi F.,
RA   Larderet G., Brahmbhatt B., Brown K., Bauce B., Muriago M., Basso C.,
RA   Thiene G., Danieli G.A., Rampazzo A.;
RT   "Identification of mutations in the cardiac ryanodine receptor gene in
RT   families affected with arrhythmogenic right ventricular
RT   cardiomyopathy type 2 (ARVD2).";
RL   Hum. Mol. Genet. 10:189-194(2001).
CC   -!- FUNCTION: Communication between transverse-tubules and
CC       sarcoplasmic reticulum. Contraction of cardiac muscle is triggered
CC       by release of calcium ions from SR following depolarization of T-
CC       tubules (By similarity).
CC   -!- SUBUNIT: Homotetramer (Potential).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
CC       produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Heart muscle, brain (cerebellum and
CC       hippocampus) and placenta.
CC   -!- DEVELOPMENTAL STAGE: Expressed in myometrium during pregnancy.
CC   -!- INDUCTION: By TGF-beta.
CC   -!- DISEASE: Defects in RYR2 are a cause of arrhythmogenic right
CC       ventricular cardiomyopathy type 2 (ARVD2). ARVD2 is an autosomal
CC       dominant disease. Characterized by partial degeneration of the
CC       myocardium of the right ventricle, electrical instability and
CC       sudden death. It is clinically defined by electrocardiographic and
CC       angiographic criteria; pathologic findings, replacement of
CC       ventricular myocardium with fatty and fibrous elements,
CC       preferentially involve the right ventricular free wall.
CC   -!- DISEASE: Defects in RYR2 are a cause of stress-induced polymorphic
CC       ventricular tachycardia (VTSIP) also known as catecholaminergic
CC       polymorphic ventricular tachycardia. VTSIP is genetic
CC       arrhythmogenic disorder characterized by stress-induced,
CC       bidirectional ventricular tachycardia that may degenerate into
CC       cardiac arrest and cause sudden death.
CC   -!- DISEASE: Defects in RYR2 are a cause of familial polymorphic
CC       ventricular tachycardia (VPFT). VPFT is an autosomal-dominant,
CC       inherited disease with a relatively early onset and a mortality
CC       rate of approximately 30% by the age of 30 years. Phenotypically,
CC       it is characterized by salvoes of bidirectional and polymorphic
CC       ventricular tachycardias in response to vigorous exercise, with no
CC       structural evidence of myocardial disease.
CC   -!- MISCELLANEOUS: The calcium release channel is modulated by calcium
CC       ions, magnesium ions, ATP and calmodulin.
CC   -!- MISCELLANEOUS: The calcium release channel activity resides in the
CC       C-terminal region while the remaining part of the protein
CC       constitutes the 'foot' structure spanning the junctional gap
CC       between the SR and the T-tubule. It is possible that the foot
CC       structure interacts with the cytoplasmic region of the
CC       dihydropyridine receptor.
CC   -!- MISCELLANEOUS: Ryanodine is an alkaloid that binds to the Ca-
CC       release channel in junctional SR and modulates its activity.
CC   -!- SIMILARITY: BELONGS TO THE RYANODINE RECEPTOR FAMILY.
CC   -!- SIMILARITY: CONTAINS 3 SPRY DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X98330; CAA66975.1; -.
DR   EMBL; Y08218; CAA69395.1; -.
DR   EMBL; X91869; CAA62975.1; -.
DR   EMBL; AJ002511; CAA05502.1; -.
DR   MIM; 180902; -.
DR   MIM; 192605; -.
DR   MIM; 600996; -.
DR   MIM; 604772; -.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001682; Channel_pore_Ca_Na.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR003608; MIR.
DR   InterPro; IPR000699; RYDR_ITPR.
DR   InterPro; IPR003032; RyR.
DR   InterPro; IPR001215; Ryanodn_receptor.
DR   InterPro; IPR003877; SPRY.
DR   InterPro; IPR003878; SPRY_domain.
DR   Pfam; PF00036; efhand; 2.
DR   Pfam; PF00520; ion_trans; 1.
DR   Pfam; PF02815; MIR; 4.
DR   Pfam; PF01365; RYDR_ITPR; 2.
DR   Pfam; PF02026; RyR; 4.
DR   Pfam; PF00622; SPRY; 3.
DR   PRINTS; PR00795; RYANODINER.
DR   SMART; SM00472; MIR; 4.
DR   SMART; SM00449; SPRY; 3.
KW   Receptor; Transmembrane; Ionic channel; Calcium channel; Repeat;
KW   Phosphorylation; Glycoprotein; Multigene family; Calmodulin-binding;
KW   Alternative splicing; Disease mutation; Polymorphism.
FT   DOMAIN        1   3089       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM   3090   3109       M' (POTENTIAL).
FT   TRANSMEM   3153   3171       M'' (POTENTIAL).
FT   TRANSMEM   3940   3959       M1 (POTENTIAL).
FT   TRANSMEM   3978   3995       M2 (POTENTIAL).
FT   TRANSMEM   4233   4256       M3 (POTENTIAL).
FT   TRANSMEM   4294   4314       M4 (POTENTIAL).
FT   TRANSMEM   4500   4520       M5 (POTENTIAL).
FT   TRANSMEM   4578   4600       M6 (POTENTIAL).
FT   TRANSMEM   4720   4740       M7 (POTENTIAL).
FT   TRANSMEM   4768   4786       M8 (POTENTIAL).
FT   TRANSMEM   4810   4827       M9 (POTENTIAL).
FT   TRANSMEM   4845   4867       M10 (POTENTIAL).
FT   DOMAIN      670    808       SPRY 1.
FT   DOMAIN     1098   1221       SPRY 2.
FT   DOMAIN     1423   1561       SPRY 3.
FT   DOMAIN     4414   4455       GLU-RICH (ACIDIC).
FT   DOMAIN      853   2925       4 X APPROXIMATE REPEATS.
FT   REPEAT      853    966       1.
FT   REPEAT      967   1080       2.
FT   REPEAT     2692   2810       3.
FT   REPEAT     2812   2925       4.
FT   BINDING    2618   3015       MODULATOR (POTENTIAL).
FT   BINDING    2774   2806       CALMODULIN (POTENTIAL).
FT   BINDING    2876   2897       CALMODULIN (POTENTIAL).
FT   BINDING    2997   3015       CALMODULIN (POTENTIAL).
FT   MOD_RES    2808   2808       PHOSPHORYLATION (BY CAM-KINASE) (BY
FT                                SIMILARITY).
FT   CARBOHYD   3427   3427       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3536   3536       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3770   3770       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3864   3864       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3865   3865       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3905   3905       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4075   4075       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4104   4104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4111   4111       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   4794   4794       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   3715   3715       E -> EVTGSQRSK (IN ISOFORM 2).
FT   VARIANT     433    433       L -> P (IN ARDV2).
FT                                /FTId=VAR_011395.
FT   VARIANT    2246   2246       S -> L (IN VTSIP).
FT                                /FTId=VAR_011396.
FT   VARIANT    2328   2328       P -> S (IN FPVT).
FT                                /FTId=VAR_011397.
FT   VARIANT    2386   2386       N -> I (IN ARDV2).
FT                                /FTId=VAR_011398.
FT   VARIANT    2474   2474       R -> S (IN VTSIP).
FT                                /FTId=VAR_011399.
FT   VARIANT    2958   2958       Q -> R.

  Query Match             13.1%;  Score 79;  DB 1;  Length 4967;
  Best Local Similarity   23.9%;  Pred. No. 1.1e+02;
  Matches   28;  Conservative   17;  Mismatches   42;  Indels   30;  Gaps    3;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL-----GQKPGGSDFLMKRL 58
           |:|||:   |||  | :  : ::|   |:|:| | | |   |     |:         |::
Db    4425 KEEEKEEKEETKSEPEKAEGEDGEKEEKAKEDKGKQKLRQLHTHRYGEPEVPESAFWKKI 4484

Qy      59 ----QKGQKYFDSGDYNM--------------------AKAKMKNKQLPSAGPDKN 90
               ||   ||    |||                      : :: |:||:    :|
Db    4485 IAYQQKLLNYFARNFYNMRMLALFVAFAINFILLFYKVSTSSVVEGKELPTRSSSEN 4541


RESULT  14
DH10_ARATH
ID   DH10_ARATH     STANDARD;      PRT;   260 AA.
AC   P42759;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Dehydrin ERD10 (Low-temperature-induced protein LTI45).
GN   ERD10 OR LTI45 OR LTI29 OR AT1G20450 OR F5M15.21.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=94348723; PubMed=8069491;
RA   Kiyosue T., Yamaguchi-Shinozaki K., Shinozaki K.;
RT   "Characterization of two cDNAs (ERD10 and ERD14) corresponding to
RT   genes that respond rapidly to dehydration stress in Arabidopsis
RT   thaliana.";
RL   Plant Cell Physiol. 35:225-231(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. LANDSBERG ERECTA; TISSUE=Leaf;
RX   MEDLINE=96046757; PubMed=7579189;
RA   Welin B.V., Olson A., Palva E.T.;
RT   "Structure and organization of two closely related low-temperature-
RT   induced dhn/lea/rab-like genes in Arabidopsis thaliana L. Heynh.";
RL   Plant Mol. Biol. 29:391-395(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
```

```
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
RN   [4]
RP   SEQUENCE OF 56-260 FROM N.A.
RC   STRAIN=CV. LANDSBERG ERECTA; TISSUE=Leaf;
RX   MEDLINE=95035985; PubMed=7948863;
RA   Welin B.V., Olson A., Nylander M., Palva E.T.;
RT   "Characterization and differential expression of dhn/lea/rab-like
RT   genes during cold acclimation and drought stress in Arabidopsis
RT   thaliana.";
RL   Plant Mol. Biol. 26:131-144(1994).
CC   -!- TISSUE SPECIFICITY: IN STEMS, CAULINE LEAVES, ROOTS AND FLOWERS.
CC       LOW LEVELS FOUND IN MATURING SEEDS. ABSENT IN DRY SEEDS.
CC   -!- INDUCTION: BY DEHYDRATION, COLD STRESS AND ABSCISIC ACID (ABA).
CC       INDUCTION BY DEHYDRATION OCCURS AFTER 1 H AND INCREASES TO A
CC       MAXIMUM AFTER 10 H. COLD STRESS INDUCTION PEAKS AT 1 H AND 5 H
CC       AFTER START OF COLD EXPOSURE.
CC   -!- SIMILARITY: BELONGS TO THE PLANT DEHYDRIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D17714; BAA04568.1; -.
DR   EMBL; X90958; CAA62448.1; -.
DR   EMBL; AC027665; AAF79613.1; -.
DR   EMBL; X77614; CAA54705.1; -.
DR   InterPro; IPR000167; Dehydrin.
DR   Pfam; PF00257; dehydrin; 1.
DR   PROSITE; PS00315; DEHYDRIN_1; 1.
DR   PROSITE; PS00823; DEHYDRIN_2; 2.
KW   Dehydrin; Repeat; Multigene family.
FT   DOMAIN       71     74       POLY-GLU.
FT   DOMAIN       88     92       POLY-GLU.
FT   DOMAIN      104    112       POLY-SER.
FT   DOMAIN      121    129       POLY-LYS.
FT   DOMAIN      183    186       POLY-GLU.
FT   DOMAIN      184    247       2 X 21 AA REPEATS, LYS-RICH.
FT   REPEAT      184    204       1.
FT   REPEAT      227    247       2.
FT   CONFLICT     56     56       H -> G (IN REF. 3).
FT   CONFLICT    144    144       Q -> E (IN REF. 3).
SQ   SEQUENCE   260 AA;  29548 MW;  318B6FC6B4A1AE5A CRC64;


  Query Match             13.0%;  Score 78.5;  DB 1;  Length 260;
  Best Local Similarity   28.4%;  Pred. No. 5.3;
  Matches   42;  Conservative   19;  Mismatches   42;  Indels   45;  Gaps   11;

Qy       2 SQKQEEENPAEETGEEKQDTQEK---EGILPERAEEA------KLKAKYPSLGQKPGGSD 52
           |   :||    | ||:|:  ::|   ||   :  ||       ::| |:| ||:|||| |
Db     109 SSSSDEEG---EDGEKKKKEKKKKIVEGDHVKTVEEENQGVMDRIKEKFP-LGEKPGGDD 164

Qy      53 FLM-------------KRLQKGQKYFDSGDYNMAKAKMKNKQLP--SAGPDKNLVTGD-- 95
             :             |  :: :| |      | | | |   || |  |: : |
Db     165 VPVVTTMPAPHSVEDHKPEEEEKKGF------MDKIKEK---LPGHSKKPEDSQVVNTTP 215

Qy      96 --HIPTP-QDLPQRKSSL---VTSKLAG 117
             ||  |:|: |      :  || |
Db     216 LVETATPIADIPEEKKGFMDKIKEKLPG 243


RESULT  15
```

```
RGS3_MOUSE
ID   RGS3_MOUSE     STANDARD;      PRT;   568 AA.
AC   Q9DC04; Q9JL22; Q9JL23;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Regulator of G-protein signaling 3 (RGS3).
GN   RGS3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (LONG ISOFORM).
RC   STRAIN=C57BL/6J; TISSUE=Lung;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [2]
RP   PARTIAL SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=20243574; PubMed=10779778;
RA   Reif K., Cyster J.G.;
RT   "RGS molecule expression in murine B lymphocytes and ability to
RT   down-regulate chemotaxis to lymphoid chemokines.";
RL   J. Immunol. 164:4720-4729(2000).
CC   -!- FUNCTION: INHIBITS SIGNAL TRANSDUCTION BY INCREASING THE GTPASE
CC       ACTIVITY OF G PROTEIN ALPHA SUBUNITS THEREBY DRIVING THEM INTO
CC       THEIR INACTIVE GDP-BOUND FORM.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: PHOSPHORYLATED BY CYCLIC GMP-DEPENDENT PROTEIN KINASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 RGS DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AK004648; BAB23439.1; -.
DR   EMBL; AF215670; AAF34627.1; -.
DR   EMBL; AF215669; AAF34626.1; -.
DR   MGD; MGI:1354734; Rgs3.
DR   InterPro; IPR000342; RGS.
DR   Pfam; PF00615; RGS; 1.
DR   PRINTS; PR01301; RGSPROTEIN.
DR   ProDom; PD001580; RGS; 1.
DR   SMART; SM00315; RGS; 1.
DR   PROSITE; PS50132; RGS; 1.
KW   Signal transduction inhibitor; Alternative splicing; Phosphorylation.
FT   DOMAIN      443    559       RGS.
```

```
FT   VARSPLIC      1    375       MISSING (IN SHORT ISOFORM).
FT   VARSPLIC    376    397       KMSGTDLTECDEASRKRKSKNI -> MLRGMYLTRNGNLQR
FT                                RHTMKE (IN SHORT ISOFORM).
SQ   SEQUENCE   568 AA;  61558 MW;  91939E22F3DA9F5C CRC64;


  Query Match             13.0%;  Score 78.5;  DB 1;  Length 568;
  Best Local Similarity   30.5%;  Pred. No. 12;
  Matches   32;  Conservative   13;  Mismatches   47;  Indels   13;  Gaps    4;

Qy      13 ETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQK-PGGSDF-------LMKRLQKGQKY 64
           ||  ::::    ||  |   | |   :| || ||: | | |         : |  ||
Db       3 ETEADEKEMPLVEGKGPGAEEPA--PSKNPSPGQELPPGQDLPPSKDPSPSQELPAGQDL 60

Qy      65 FDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSS 109
             | | : ::     : ||   | |:      :|  |||| || |
Db      61 PPSKDPSPSQELPAGQDLP---PSKDPSPSQELPVGQDLPPRKDS 102


RESULT  16
ATRX_CAEEL
ID   ATRX_CAEEL     STANDARD;      PRT;  1359 AA.
AC   Q9U7E0; O02061;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
GN   XNP-1 OR B0041.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=99365296; PubMed=10433961;
RA   Villard L., Fontes M., Ewbank J.J.;
RT   "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RT   the human XNP/ATR-X gene.";
RL   Gene 236:13-19(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Fulton R., Wohldmann P.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC       GENE EXPRESSION BY AFFECTING CHROMATIN (Potential).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF134186; AAD55361.1; -.
DR   EMBL; AF000196; AAC24256.1; -.
DR   WormPep; B0041.7; CE17314.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding.
FT   NP_BIND     496    503       ATP (POTENTIAL).
FT   SITE        636    639       DEAH BOX.
FT   DOMAIN       67     70       POLY-ASP.
FT   DOMAIN      266    272       POLY-GLU.
FT   DOMAIN      276    281       POLY-LYS.
FT   DOMAIN      372    375       POLY-LYS.
FT   DOMAIN      603    608       POLY-LYS.
FT   DOMAIN      859    862       POLY-LYS.
FT   CONFLICT    479    479       C -> F (IN REF. 2).
SQ   SEQUENCE   1359 AA;  156191 MW;  EB4342547D4F4E64 CRC64;


  Query Match             13.0%;  Score 78.5;  DB 1;  Length 1359;
  Best Local Similarity   26.9%;  Pred. No. 31;
  Matches   32;  Conservative   24;  Mismatches   48;  Indels   15;  Gaps    5;

Qy       2 SQKQEEENPAEETGEE---KQDTQEKEGILPERAE--EAKLKAKYPSLGQKPGGSDFLMK 56
           ::||     :||: ||   |:  : ||  : :|||  |   : : ||   | |     :|
Db     147 TKKQTSSESSEESEEERKVKKSKKNKEKSVKKRAETSEESDEDEKPSKKSKKG-----LK 201

Qy      57 RLQKGQKYFDSGD-YNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSK 114
           :  | :   :| |   : |:| |:|::      |   : |  |  :   :|| |  :|:
Db     202 KKAKSESESESEDEKEVKKSKKKSKKVVK----KESESEDEAPEKKKTEKRKRSKTSSE 256


RESULT  17
IF2P_HUMAN
ID   IF2P_HUMAN     STANDARD;      PRT;  1220 AA.
AC   O60841; O95805; Q9UF81; Q9UMN7;
DT   15-DEC-1998 (Rel. 37, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Translation initiation factor IF-2.
GN   IF2 OR KIAA0741.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99087487; PubMed=9872452;
RA   Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XI.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 5:277-286(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cervical carcinoma;
RX   MEDLINE=99362399; PubMed=10432305;
RA   Wilson S.A., Sieiro-Vazquez C., Edwards N.J., Iourin O., Byles E.D.,
RA   Kotsopoulou E., Adamson C.S., Kingsman S.M., Kingsman A.J.,
RA   Martin-Rendon E.;
RT   "Cloning and characterization of hIF2, a human homologue of bacterial
RT   translation initiation factor 2 and its interaction with HIV-1
RT   matrix.";
RL   Biochem. J. 342:97-103(1999).
RN   [3]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   TISSUE=Testis;
RX   MEDLINE=99218282; PubMed=10200264;
RA   Lee J.H., Choi S.K., Roll-Mecak A., Burley S.K., Dever T.E.;
RT   "Universal conservation in translation initiation revealed by human
RT   and archaeal homologs of bacterial translation initiation factor
RT   IF2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:4342-4347(1999).
RN   [4]
RP   SEQUENCE OF 89-1220 FROM N.A.
RC   TISSUE=Testis;
RA   Koehrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 833-1220 FROM N.A.
RX   PubMed=11124703;
RA   Stanchi F., Bertocco E., Toppo S., Dioguardi R., Simionati B.,
RA   Cannata N., Zimbello R., Lanfranchi G., Valle G.;
```

```
RT   "Characterization of 16 novel human genes showing high similarity to
RT   yeast sequences.";
RL   Yeast 18:69-80(2001).
CC   -!- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING
CC       THE BINDING OF THE FORMYLMETHIONINE-TRNA TO RIBOSOMES. SEEMS TO
CC       FUNCTION ALONG WITH EIF-2 (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB018284; BAA34461.1; -.
DR   EMBL; AJ006776; CAB44357.1; -.
DR   EMBL; AF078035; AAD16006.1; -.
DR   EMBL; AL133563; CAB63717.1; -.
DR   EMBL; AJ006412; CAA07018.1; -.
DR   MIM; 606086; -.
DR   InterPro; IPR000795; GTP_EFTU.
DR   InterPro; IPR004161; GTP_EFTU_D2.
DR   InterPro; IPR000178; IF2.
DR   Pfam; PF00009; GTP_EFTU; 1.
DR   Pfam; PF03144; GTP_EFTU_D2; 2.
DR   PRINTS; PR00315; ELONGATNFCT.
DR   PROSITE; PS01176; IF2; FALSE_NEG.
KW   Initiation factor; Protein biosynthesis; GTP-binding.
FT   NP_BIND     638    645       GTP (BY SIMILARITY).
FT   DOMAIN       39     50       POLY-LYS.
FT   DOMAIN       94     99       POLY-LYS.
FT   DOMAIN      138    142       POLY-ASP.
FT   DOMAIN      313    322       POLY-LYS.
FT   DOMAIN      353    356       POLY-GLU.
FT   DOMAIN      361    364       POLY-GLU.
FT   DOMAIN      491    496       POLY-GLU.
FT   DOMAIN      529    567       ASP/GLU-RICH (ACIDIC).
FT   MUTAGEN     640    640       V->G: LOSS OF ACTIVITY IN VIVO. RETAINS
FT                                FULL ACTIVITY IN VITRO.
FT   MUTAGEN     706    706       H->E: LOSS OF ACTIVITY; BOTH IN VIVO AND
FT                                IN VITRO.
FT   MUTAGEN     706    706       H->Q: LOSS OF ACTIVITY IN VIVO. PARTIAL
FT                                ACTIVITY IN VITRO.
FT   MUTAGEN     759    759       D->N: LOSS OF ACTIVITY; BOTH IN VIVO AND
FT                                IN VITRO.
FT   CONFLICT     64     64       E -> G (IN REF. 1).
FT   CONFLICT     92     92       T -> I (IN REF. 2).
FT   CONFLICT    180    180       I -> M (IN REF. 3).
FT   CONFLICT    256    256       K -> R (IN REF. 3).
FT   CONFLICT    522    522       T -> K (IN REF. 1).
FT   CONFLICT    549    549       E -> V (IN REF. 3).
FT   CONFLICT    669    669       G -> W (IN REF. 3).
FT   CONFLICT    894    894       E -> K (IN REF. 2).
SQ   SEQUENCE   1220 AA;  138799 MW;  010CC15DE475E5B8 CRC64;


  Query Match             12.7%;  Score 77;  DB 1;  Length 1220;
  Best Local Similarity   26.2%;  Pred. No. 37;
  Matches   28;  Conservative   21;  Mismatches   30;  Indels   28;  Gaps    6;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGIL--------PERAEEAKLKAKYPSLGQKPGGSDFL 54
           :::|||   ||  || :: :|:||           |:  : |   |  :| :||
Db     532 EEEEEEEEEEEEDEESEEEEEEEGESEGSEGDEEDEKVSDEKDSGK--TLDKKPS----- 584

Qy      55 MKRLQKGQKYFDSGD--------YNMAKAKMKNKQLPSAGPDKNLVT 93
             | :    :| || |        |: || ::: ::|  :   ||: |
Db     585 -KEMSSDSEY-DSDDDRTKEERAYDKAKRRIEKRRLEHS---KNVNT 626


RESULT  18
CALD_MELGA
ID   CALD_MELGA     STANDARD;      PRT;   239 AA.
AC   P13505;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Caldesmon, smooth muscle (CDM) (Fragment).
GN   CALD1 OR CAD.
OS   Meleagris gallopavo (Common turkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Meleagrididae; Meleagris.
OX   NCBI_TaxID=9103;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Gizzard;
RX   MEDLINE=89273603; PubMed=2730648;
RA   Leszyk J., Mornet D., Audemard E., Collins J.H.;
RT   "Caldesmon structure and function: sequence analysis of a 35
RT   kilodalton actin- and calmodulin-binding fragment from the C-terminus
RT   of the turkey gizzard protein.";
RL   Biochem. Biophys. Res. Commun. 160:1371-1378(1989).
RN   [2]
RP   SEQUENCE OF 1-96.
RC   TISSUE=Gizzard;
RX   MEDLINE=89228015; PubMed=2653315;
RA   Leszyk J., Mornet D., Audemard E., Collins J.H.;
RT   "Amino acid sequence of a 15 kilodalton actin-binding fragment of
RT   turkey gizzard caldesmon: similarity with dystrophin, tropomyosin and
RT   the tropomyosin-binding region of troponin T.";
RL   Biochem. Biophys. Res. Commun. 160:210-216(1989).
CC   -!- FUNCTION: CONTROL OF ACTOMYOSIN INTERACTIONS IN SMOOTH MUSCLE
CC       AND NONMUSCLE CELLS (COULD ACT AS A BRIDGE BETWEEN MYOSIN AND
CC       ACTIN FILAMENTS). INHIBITS THE ACTIN-ACTIVATED ATPASE OF MYOSIN
CC       THIS INHIBITION IS ATTENUATED BY CALCIUM-CALMODULIN AND IS
CC       POTENTIATED BY TROPOMYOSIN. INTERACTS WITH ACTIN, MYOSIN,
CC       2 MOLECULES OF TROPOMYOSIN AND WITH CALMODULIN.
CC   -!- SUBCELLULAR LOCATION: ON THIN FILAMENTS IN SMOOTH MUSCLE AND ON
CC       STRESS FIBERS IN FIBROBLASTS (NONMUSCLE).
DR   PIR; A32488; A32488.
DR   PIR; A32420; A32420.
KW   Muscle protein; Actin-binding; Calmodulin-binding.
FT   NON_TER       1      1
FT   NON_TER     239    239
SQ   SEQUENCE   239 AA;  26561 MW;  D395B8BB465B8895 CRC64;


  Query Match             12.6%;  Score 76.5;  DB 1;  Length 239;
  Best Local Similarity   29.2%;  Pred. No. 7.1;
  Matches   38;  Conservative   13;  Mismatches   40;  Indels   39;  Gaps    7;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLG----QKP-------GGS 51
           ||:::|    :  ||::  : || |   ||| |: : | |  |    :||       | |
Db      46 QKKKQEEAERKIREEEEKKRMKEEIERRRAEAAEKRQKMPEDGVSEEKKPFKCFSPKGSS 105

Qy      52 -------DFLMKRLQK-GQK---------YFDSGDYNMAKAKMKNKQLPSAGPDKNLVTG 94
                :|| |  || | |           ||       | : ||   :| | |
Db     106 LKIEERAEFLNKSAQKSGMKPAHTTAVVSKIDSRLEQYTSAVVGNK---AAKPAK----- 157

Qy      95 DHIPTPQDLP 104
             |   |||
Db     158 ---PAASDLP 164


RESULT  19
CALD_CHICK
ID   CALD_CHICK     STANDARD;      PRT;   771 AA.
AC   P12957; Q90756; Q90761; Q92018; Q99230; Q03698;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Caldesmon (CDM).
GN   CALD1 OR CAD.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (GIZZARD H-CAD ISOFORM).
RC   TISSUE=Gizzard, and Oviduct;
RX   MEDLINE=89340480; PubMed=2760048;
RA   Bryan J., Imai M., Lee R., Moore P., Cook R.G., Lin W.-G.;
RT   "Cloning and expression of a smooth muscle caldesmon.";
RL   J. Biol. Chem. 264:13873-13879(1989).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (GIZZARD H-CAD ISOFORM).
RC   TISSUE=Gizzard;
RX   MEDLINE=90026426; PubMed=2803315;
RA   Hayashi K., Kanda K., Kimizuka F., Kato I., Sobue K.;
RT   "Primary structure and functional expression of h-caldesmon
RT   complementary DNA.";
RL   Biochem. Biophys. Res. Commun. 164:503-511(1989).
RN   [3]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (BRAIN L-CAD ISOFORM).
RC   TISSUE=Brain;
RX   MEDLINE=91093148; PubMed=1824698;
RA   Hayashi K., Fujio Y., Kato I., Sobue K.;
RT   "Structural and functional relationships between h- and
RT   l-caldesmons.";
RL   J. Biol. Chem. 266:355-361(1991).
RN   [4]
RP   SEQUENCE FROM N.A. (GIZZARD H-CAD; BRAIN L-CAD AND GIZZARD L-CAD).
RX   MEDLINE=94071934; PubMed=8250919;
RA   Haruna M., Hayashi K., Yano H., Takeuchi O., Sobue K.;
RT   "Common structural and expressional properties of vertebrate
RT   caldesmon genes.";
RL   Biochem. Biophys. Res. Commun. 197:145-153(1993).
RN   [5]
RP   SEQUENCE FROM N.A. (GIZZARD L-CAD ISOFORM).
RC   TISSUE=Gizzard;
RX   MEDLINE=92042686; PubMed=1939602;
RA   Bryan J., Lee R.;
RT   "Sequence of an avian non-muscle caldesmon.";
RL   J. Muscle Res. Cell Motil. 12:372-375(1991).
RN   [6]
RP   SEQUENCE OF 1-15 FROM N.A. (BRAIN L-CAD ISOFORM).
RC   TISSUE=Brain;
RX   MEDLINE=94271210; PubMed=8002994;
RA   Yano H., Hayashi K., Haruna M., Sobue K.;
RT   "Identification of two distinct promoters in the chicken caldesmon
RT   gene.";
RL   Biochem. Biophys. Res. Commun. 201:618-626(1994).
RN   [7]
RP   SEQUENCE OF 451-756 FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Gizzard;
RX   MEDLINE=89273666; PubMed=2730665;
RA   Hayashi K., Yamada S., Kanda K., Kimizuka F., Kato I., Sobue K.;
RT   "35 kDa fragment of h-caldesmon conserves two consensus sequences of
RT   the tropomyosin-binding domain in troponin T.";
RL   Biochem. Biophys. Res. Commun. 161:38-45(1989).
RN   [8]
RP   SEQUENCE OF 498-525.
RX   MEDLINE=88293484; PubMed=3401222;
RA   Mornet D., Audemard E., Derancourt J.;
RT   "Identification of a 15 kilodalton actin binding region on gizzard
RT   caldesmon probed by chemical cross-linking.";
RL   Biochem. Biophys. Res. Commun. 154:564-571(1988).
RN   [9]
RP   PHOSPHORYLATION SITES.
RX   MEDLINE=92041815; PubMed=1939059;
RA   Mak A.S., Carpenter M., Smillie L.B., Wang J.H.;
RT   "Phosphorylation of caldesmon by p34cdc2 kinase. Identification of
RT   phosphorylation sites.";
RL   J. Biol. Chem. 266:19971-19975(1991).
CC   -!- FUNCTION: ACTIN- AND MYOSIN-BINDING PROTEIN IMPLICATED IN THE
CC       REGULATION OF ACTOMYOSIN INTERACTIONS IN SMOOTH MUSCLE AND
CC       NONMUSCLE CELLS (COULD ACT AS A BRIDGE BETWEEN MYOSIN AND ACTIN
CC       FILAMENTS). STIMULATES ACTIN BINDING OF TROPOMYOSIN WHICH
CC       INCREASES THE STABILIZATION OF ACTIN FILAMENT STRUCTURE. IN MUSCLE
CC       TISSUES, INHIBITS THE ACTOMYOSIN ATPASE BY BINDING TO F-ACTIN.
CC       THIS INHIBITION IS ATTENUATED BY CALCIUM-CALMODULIN AND IS
CC       POTENTIATED BY TROPOMYOSIN. INTERACTS WITH ACTIN, MYOSIN, TWO
CC       MOLECULES OF TROPOMYOSIN AND WITH CALMODULIN. ALSO PLAY AN
CC       ESSENTIAL ROLE DURING CELLULAR MITOSIS AND RECEPTOR CAPPING.
CC   -!- SUBCELLULAR LOCATION: ON THIN FILAMENTS IN SMOOTH MUSCLE AND ON
CC       STRESS FIBERS IN FIBROBLASTS (NONMUSCLE) (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; GIZZARD H-CAD (SHOWN HERE),
CC       BRAIN L-CAD AND GIZZARD L-CAD; ARE PRODUCED BY ALTERNATIVE
CC       SPLICING.
CC   -!- TISSUE SPECIFICITY: HIGH-MOLECULAR-WEIGHT CALDESMON (H-CALDESMON)
CC       IS PREDOMINANTLY EXPRESSED IN SMOOTH MUSCLES, WHEREAS LOW-
CC       MOLECULAR-WEIGHT CALDESMON (L-CALDESMON) IS WIDELY DISTRIBUTED IN
CC       NON-MUSCLE TISSUES AND CELLS. NOT EXPRESSED IN SKELETAL MUSCLE OR
CC       HEART.
CC   -!- DOMAIN: THE N-TERMINAL PART SEEMS TO BE A MYOSIN/CALMODULIN-
CC       BINDING DOMAIN, AND THE C-TERMINAL A TROPOMYOSIN/ACTIN/CALMODULIN-
CC       BINDING DOMAIN. THESE TWO DOMAINS ARE SEPARATED BY A CENTRAL
CC       HELICAL REGION IN THE MUSCLE FORMS.
CC   -!- PTM: IN NON-MUSCLE CELLS, PHOSPHORYLATION BY CDC2 DURING MITOSIS
CC       CAUSES CALDESMON TO DISSOCIATE FROM MICROFILAMENTS.
CC       PHOSPHORYLATION REDUCES CALDESMON BINDING TO ACTIN, MYOSIN, AND
CC       CALMODULIN AS WELL AS ITS INHIBITION OF ACTOMYOSIN ATPASE
CC       ACTIVITY. PHOSPHORYLATION ALSO OCCURS IN BOTH QUIESCENT AND
CC       DIVIDING SMOOTH MUSCLE CELLS WITH SIMILAR EFFECTS ON THE
CC       INTERACTION WITH ACTIN AND CALMODULIN AND ON MICROFILAMENTS
CC       REORGANIZATION (BY SIMILARITY).
CC   -!- SIMILARITY: TO A TROPOMYOSIN BINDING SITE DOMAIN OF TROPONIN T.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04968; AAA49067.1; -.
DR   EMBL; D17648; BAA04539.1; -.
DR   EMBL; D17634; BAA04539.1; JOINED.
DR   EMBL; D17636; BAA04539.1; JOINED.
DR   EMBL; D17637; BAA04539.1; JOINED.
DR   EMBL; D17638; BAA04539.1; JOINED.
DR   EMBL; D17639; BAA04539.1; JOINED.
DR   EMBL; D17640; BAA04539.1; JOINED.
DR   EMBL; D17641; BAA04539.1; JOINED.
DR   EMBL; D17642; BAA04539.1; JOINED.
DR   EMBL; D17643; BAA04539.1; JOINED.
DR   EMBL; D17644; BAA04539.1; JOINED.
DR   EMBL; D17645; BAA04539.1; JOINED.
DR   EMBL; D17646; BAA04539.1; JOINED.
DR   EMBL; M28417; AAA48810.1; -.
DR   EMBL; M60620; AAA48936.1; -.
DR   EMBL; D17648; BAA04538.1; -.
DR   EMBL; D17635; BAA04538.1; JOINED.
DR   EMBL; D17636; BAA04538.1; JOINED.
DR   EMBL; D17637; BAA04538.1; JOINED.
DR   EMBL; D17639; BAA04538.1; JOINED.
DR   EMBL; D17640; BAA04538.1; JOINED.
DR   EMBL; D17641; BAA04538.1; JOINED.
DR   EMBL; D17642; BAA04538.1; JOINED.
DR   EMBL; D17643; BAA04538.1; JOINED.
DR   EMBL; D17644; BAA04538.1; JOINED.
DR   EMBL; D17645; BAA04538.1; JOINED.
DR   EMBL; D17646; BAA04538.1; JOINED.
DR   EMBL; D17647; BAA04538.1; JOINED.
DR   EMBL; D17648; BAA04540.1; -.
DR   EMBL; D17634; BAA04540.1; JOINED.
DR   EMBL; D17636; BAA04540.1; JOINED.
DR   EMBL; D17637; BAA04540.1; JOINED.
DR   EMBL; D17639; BAA04540.1; JOINED.
```

```
DR   EMBL; D17640; BAA04540.1; JOINED.
DR   EMBL; D17641; BAA04540.1; JOINED.
DR   EMBL; D17642; BAA04540.1; JOINED.
DR   EMBL; D17643; BAA04540.1; JOINED.
DR   EMBL; D17644; BAA04540.1; JOINED.
DR   EMBL; D17645; BAA04540.1; JOINED.
DR   EMBL; D17646; BAA04540.1; JOINED.
DR   EMBL; D17647; BAA04540.1; JOINED.
DR   EMBL; M59762; AAA48649.1; -.
DR   EMBL; D17552; BAA04490.1; -.
DR   EMBL; M26684; AAA48811.1; -.
DR   PIR; A32642; A32642.
DR   HSSP; P05412; 1FOS.
DR   InterPro; IPR000075; Caldesmon.
DR   Pfam; PF02029; Caldesmon; 1.
DR   PRINTS; PR01076; CALDESMON.
KW   Muscle protein; Actin-binding; Calmodulin-binding; Phosphorylation;
KW   Repeat; Alternative splicing.
FT   DOMAIN       26    199       MYOSIN AND CALMODULIN-BINDING.
FT   DOMAIN      251    390       10 X 13 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT      251    265       1.
FT   REPEAT      266    278       2.
FT   REPEAT      279    291       3.
FT   REPEAT      294    306       4.
FT   REPEAT      309    321       5.
FT   REPEAT      324    336       6.
FT   REPEAT      337    349       7.
FT   REPEAT      350    362       8.
FT   REPEAT      363    375       9.
FT   REPEAT      378    390       10.
FT   DOMAIN      523    580       TROPOMYOSIN-BINDING (POTENTIAL).
FT   DOMAIN      622    632       TROPOMYOSIN-BINDING (POTENTIAL).
FT   DOMAIN      612    644       STRONG ACTIN-BINDING.
FT   DOMAIN      674    680       CALMODULIN-BINDING.
FT   DOMAIN      726    752       WEAK ACTIN-BINDING.
FT   DOMAIN       39     46       POLY-ARG.
FT   DOMAIN      539    542       POLY-GLU.
FT   DOMAIN      556    559       POLY-GLU.
FT   MOD_RES     485    485       PHOSPHORYLATION (POTENTIAL).
FT   MOD_RES     597    597       PHOSPHORYLATION (BY CDC2).
FT   MOD_RES     682    682       PHOSPHORYLATION (BY CDC2).
FT   MOD_RES     688    688       PHOSPHORYLATION (BY CDC2).
FT   MOD_RES     711    711       PHOSPHORYLATION (BY CDC2).
FT   MOD_RES     717    717       PHOSPHORYLATION (BY CDC2).
FT   VARSPLIC      1     24       MDDFERRRELRRQKREEMRLEAER -> MISRSYCRQNLS

  Query Match             12.6%;  Score 76.5;  DB 1;  Length 771;
  Best Local Similarity   29.2%;  Pred. No. 25;
  Matches   38;  Conservative   13;  Mismatches   40;  Indels   39;  Gaps    7;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLG----QKP-------GGS 51
           ||:::|    :  ||:  : || |   ||| |: : | |  |    :||       | |
Db     543 QKKKQEEAERKIREEEEKKRMKEEIERRRAEAAEKRQKVPEDGVSEEKKPFKCFSPKGSS 602

Qy      52 -------DFLMKRLQK-GQK---------YFDSGDYNMAKAKMKNKQLPSAGPDKNLVTG 94
                :|| |  || | |           ||       | : ||   :| | |
Db     603 LKIEERAEFLNKSAQKSGMKPAHTTAVVSKIDSRLEQYTSAVVGNK---AAKPAK----- 654

Qy      95 DHIPTPQDLP 104
              |   |||
Db     655 ---PAASDLP 661


RESULT  20
MOZ_HUMAN
ID   MOZ_HUMAN      STANDARD;      PRT;  2004 AA.
AC   Q92794;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Monocytic leukemia zinc finger protein (Zinc finger protein 220).
GN   ZNF220 OR MOZ.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96376968; PubMed=8782817;
RA   Borrow J., Stanton V.P. Jr., Andresen J.M., Becher R., Behm F.G.,
RA   Chaganti R.S.K., Civin C.I., Disteche C., Dube I., Frischauf A.M.,
RA   Horsman D., Mitelman F., Volinia S., Watmore A.E., Housman D.E.;
RT   "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT   a putative acetyltransferase to the CREB-binding protein.";
RL   Nat. Genet. 14:33-41(1996).
CC   -!- FUNCTION: MAY REPRESENT A CHROMATIN-ASSOCIATED ACETYLTRANSFERASE.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DISEASE: PARTICIPATES IN A T(8;16)(P11;P13) CHROMOSOMAL
CC       TRANSLOCATION THAT PRODUCES A MOZ-CBP CHIMAERA OBSERVED IN THE
CC       M4/M5 SUBTYPE OF ACUTE MYELOID LEUKEMIA.
CC   -!- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
CC   -!- SIMILARITY: BELONGS TO THE MYST (SAS/MOZ) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U47742; AAC50662.1; -.
DR   HSSP; Q60631; 4GBQ.
DR   MIM; 601408; -.
DR   InterPro; IPR001386; Linker_histone.
DR   InterPro; IPR002717; MOZ_SAS.
DR   InterPro; IPR001965; PHD.
DR   Pfam; PF01853; MOZ_SAS; 1.
DR   Pfam; PF00628; PHD; 2.
DR   SMART; SM00526; H15; 1.
DR   SMART; SM00249; PHD; 2.
KW   Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;
KW   Nuclear protein.
FT   ZN_FING      206    256       PHD-TYPE 1.
FT   ZN_FING      259    313       PHD-TYPE 2.
FT   DOMAIN       371    379       POLY-SER.
FT   ZN_FING      538    560       C2HC-TYPE.
FT   DOMAIN       788    801       POLY-GLU.
FT   DOMAIN       989    995       POLY-GLU.
FT   DOMAIN      1019   1026       POLY-ARG.
FT   DOMAIN      1069   1078       POLY-GLU.
FT   DOMAIN      1147   1150       POLY-LYS.
FT   DOMAIN      1221   1242       GLU-RICH.
FT   DOMAIN      1267   1302       GLU-RICH.
FT   DOMAIN      1411   1414       POLY-GLU.
FT   DOMAIN      1593   1597       POLY-SER.
FT   DOMAIN      1643   1704       GLN/PRO-RICH.
FT   DOMAIN      1897   1977       MET-RICH.
FT   SITE        1546   1547       BREAKPOINT FOR TRANSLOCATION TO FORM
FT                                 MOZ-CBP.
SQ   SEQUENCE   2004 AA;  225054 MW;  9FFBBAC3792854BA CRC64;


  Query Match             12.6%;  Score 76.5;  DB 1;  Length 2004;
  Best Local Similarity   20.7%;  Pred. No. 70;
  Matches   28;  Conservative   28;  Mismatches   44;  Indels   35;  Gaps    4;

Qy       4 KQEEENP--AEETGEEKQDTQEKEGILPERAEEAKLK------------------AKYP 42
             ::|||  |  :||  : ::: || |   ||  |:|  :                  |
Db    1271 EEEEEKPRVSEEQRQSEEEQQELEEPEPEEEEDAAAETAQNDDHDADDEDDGHLESTKKK 1330

Qy      43 SLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD 102
             | ::|   |  :|     |: |   :   :: |:|:|:                : : ::
Db    1331 ELEEQPTRED--VKEEPGVQESFLDANMQKSREKIKDKE------------ETELDSEEE 1376

Qy     103 LPQRKSSLVTSKLAG 117
```

```
          |   :|:|: ::||
Db   1377 QPSHDTSVVSEQMAG 1391


RESULT  21
TRT2_RABIT
ID   TRT2_RABIT     STANDARD;      PRT;   300 AA.
AC   P09741; P09742;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Troponin T, cardiac muscle isoforms (TnTC).
GN   TNNT2.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=94084880; PubMed=8261593;
RA   Greig A., Hirschberg Y., Anderson P.A.W., Hainsworth C.,
RA   Malouf N.N., Oakeley A.E., Kay B.K.;
RT   "Molecular basis of cardiac troponin T isoform heterogeneity in
RT   rabbit heart.";
RL   Circ. Res. 74:41-47(1994).
RN   [2]
RP   SEQUENCE OF 1-15; 22; 33-206 AND 216-300.
RX   MEDLINE=87057385; PubMed=3782144;
RA   Pearlstone J.R., Carpenter M.R., Smillie L.B.;
RT   "Amino acid sequence of rabbit cardiac troponin T.";
RL   J. Biol. Chem. 261:16795-16810(1986).
CC   -!- FUNCTION: TROPONIN T IS THE TROPOMYOSIN-BINDING SUBUNIT OF
CC       TROPONIN, THE THIN FILAMENT REGULATORY COMPLEX WHICH CONFERS
CC       CALCIUM-SENSITIVITY TO STRIATED MUSCLE ACTOMYOSIN ATPASE ACTIVITY.
CC   -!- SUBUNIT: BINDS WITH TROPONINS I AND C TO MAKE THE THIN-FILAMENT
CC       REGULATORY COMPLEX, TROPONIN.
CC   -!- ALTERNATIVE PRODUCTS: THE DIFFERENT ISOFORMS ARE GENERATED BY A
CC       DIFFERENTIAL MRNA SPLICING MECHANISM. MYOFILAMENTS CONTAINING A
CC       RELATIVELY GREATER AMOUNT OF A HIGH MOLECULAR WEIGHT CTNT ISOFORM
CC       BOUND LESS CALCIUM TO MYOFIBRILLAR TNC THAN MYOFILAMENTS THAT DID
CC       NOT CONTAIN THE HIGH MOLECULAR WEIGHT CTNT. THE MAJOR ISOFORM IN
CC       ADULT HEART IS CTNT4.
CC   -!- MISCELLANEOUS: THE SEQUENCE OF ISOFORM CTNT1 IS SHOWN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L40178; AAB51160.1; ALT_INIT.
DR   PIR; A25345; A25345.
DR   PIR; B25345; B25345.
DR   InterPro; IPR001978; Troponin.
DR   Pfam; PF00992; Troponin; 1.
KW   Muscle protein; Alternative splicing; Multigene family;
KW   Phosphorylation.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       PHOSPHORYLATION (BY CK2)
FT                                (BY SIMILARITY).
FT   VARSPLIC     16     21       MISSING (IN ISOFORM CTNT2 AND ISOFORM
FT                                CTNT4).
FT   VARSPLIC     23     32       MISSING (IN ISOFORM CTNT3 AND ISOFORM
FT                                CTNT4).
FT   CONFLICT     45     45       R -> A (IN REF. 2).
FT   CONFLICT    138    138       D -> E (IN REF. 2).
FT   CONFLICT    146    146       D -> ER (IN REF. 2).
FT   CONFLICT    183    183       E -> D (IN REF. 2).
FT   CONFLICT    194    194       MISSING (IN REF. 1).
SQ   SEQUENCE   300 AA;  35745 MW;  DA97C09EA37169A5 CRC64;
```

```
  Query Match             12.5%;  Score 75.5;  DB 1;  Length 300;
  Best Local Similarity   29.4%;  Pred. No. 11;
  Matches   20;  Conservative   17;  Mismatches   18;  Indels   13;  Gaps    3;

Qy       2 SQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAK--YPSL-------GQKPGGSD 52
           :: : ||  ||| |:|::| :::::|      ||:| | :   |:|       |::    |
Db      47 AEAETEETQAEEDGQEEEDKEDEDG----PVEESKPKPRPFMPNLVPPKIPDGERVDFDD 102

Qy      53 FLMKRLQK 60
             ||::|
Db     103 IHRKRMEK 110


RESULT  22
FKB3_YEAST
ID   FKB3_YEAST     STANDARD;      PRT;   411 AA.
AC   P38911;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   FK506-binding nuclear protein (Peptidyl-prolyl cis-trans isomerase)
DE   (PPiase) (EC 5.2.1.8) (Proline rotamase) (Nucleolar proline
DE   isomerase) (FKBP-70).
GN   FPR3 OR NPI46 OR YML074C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YNN 214;
RX   MEDLINE=95050937; PubMed=7525596;
RA   Benton B.M., Zang J.-H., Thorner J.;
RT   "A novel FK506- and rapamycin-binding protein (FPR3 gene product) in
RT   the yeast Saccharomyces cerevisiae is a proline rotamase localized to
RT   the nucleolus.";
RL   J. Cell Biol. 127:623-639(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=94327703; PubMed=8051210;
RA   Shan X., Xue Z., Melese T.;
RT   "Yeast NPI46 encodes a novel prolyl cis-trans isomerase that is
RT   located in the nucleolus.";
RL   J. Cell Biol. 126:853-862(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95010686; PubMed=7925954;
RA   Manning-Krieg U.C., Henriquez R., Cammas F., Graff P.,
RA   Gaveriaux S., Movva N.R.;
RT   "Purification of FKBP-70, a novel immunophilin from Saccharomyces
RT   cerevisiae, and cloning of its structural gene, FPR3.";
RL   FEBS Lett. 352:98-103(1994).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Brown D., Bowman S., Barrell B.G., Rajandream M.A.;
RL   Submitted (OCT-1994) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   PHOSPHORYLATION OF TYR-184 AND SER-186.
RX   MEDLINE=97294697; PubMed=9148902;
RA   Wilson L.K., Dhillon N., Thorner J., Martin G.S.;
RT   "Casein kinase II catalyzes tyrosine phosphorylation of the yeast
RT   nucleolar immunophilin Fpr3.";
RL   J. Biol. Chem. 272:12961-12967(1997).
CC   -!- FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS. FK506- AND
CC       RAPAMYCIN-BINDING PROTEIN. SPECIFICALLY BINDS NUCLEAR LOCALIZATION
CC       SEQUENCES. MAY BE INVOLVED IN THE ASSEMBLY OR FOLDING OF RIBOSOMAL
CC       PROTEINS.
CC   -!- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC       PEPTIDE BONDS IN OLIGOPEPTIDES.
```

```
CC   -!- ENZYME REGULATION: INHIBITED BY BOTH FK506 AND RAPAMYCIN.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
CC   -!- PTM: PHOSPHORYLATED AT TYROSINE AND DEPHOSPHORYLATED BY THE
CC       PHOSPHOTYROSINE-SPECIFIC PHOSPHOPROTEIN PHOSPHATASE PTP1.
CC   -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L34569; AAB04165.1; -.
DR   EMBL; X79379; CAA55924.1; -.
DR   EMBL; S73876; AAB31995.1; -.
DR   EMBL; Z46373; CAA86504.1; -.
DR   PIR; S47927; S47927.
DR   PIR; A53544; A53544.
DR   HSSP; P20081; 1YAT.
DR   SGD; S0004539; NPI46.
DR   InterPro; IPR001179; FKBP_PPIase.
DR   Pfam; PF00254; FKBP; 1.
DR   PROSITE; PS00453; FKBP_PPIASE_1; 1.
DR   PROSITE; PS00454; FKBP_PPIASE_2; 1.
DR   PROSITE; PS50059; FKBP_PPIASE_3; 1.
KW   Isomerase; Rotamase; Nuclear protein; Phosphorylation.
FT   DOMAIN       60     87       ASP/GLU-RICH (HIGHLY ACIDIC).
FT   DOMAIN       88     99       LYS-RICH (HIGHLY BASIC).
FT   DOMAIN      101    119       ASP/GLU-RICH (HIGHLY ACIDIC).
FT   DOMAIN      173    248       ASP/GLU-RICH (HIGHLY ACIDIC).
FT   DOMAIN      250    298       LYS-RICH (HIGHLY BASIC).
FT   DOMAIN      256    271       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      324    411       PPIASE, FKBP-TYPE.
FT   MOD_RES     184    184       PHOSPHORYLATION (BY CK2).
FT   MOD_RES     186    186       PHOSPHORYLATION (BY CK2).
FT   CONFLICT    122    122       L -> P (IN REF. 1).
FT   CONFLICT    240    240       E -> EEE (IN REF. 1).
FT   CONFLICT    335    335       L -> F (IN REF. 1).
SQ   SEQUENCE   411 AA;  46553 MW;  A01D24DE0078FE11 CRC64;


  Query Match             12.5%;  Score 75.5;  DB 1;  Length 411;
  Best Local Similarity   25.2%;  Pred. No. 16;
  Matches   28;  Conservative   20;  Mismatches   52;  Indels   11;  Gaps    2;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           : :: ||:|  ||  ||:::: ::|| : ||  :  | | :     ::        |: :|
Db     219 VQEEDEEDNDGEEEQEEEEEEEEQKEEVKPEPKKSKKEKKRKHEEKEEE-------KKAKK 271

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHI----PTPQDLPQRK 107
            :|     |      | |:|:       |: |    |     |  | || |
Db     272 VKKVEFKKDLEEGPTKPKSKKEQDKHKPKSKVLEGGIVIEDRTIGDGPQAK 322


RESULT  23
SMY2_YEAST
ID   SMY2_YEAST     STANDARD;      PRT;   790 AA.
AC   P32909;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   SMY2 protein.
GN   SMY2 OR YBR172C OR YBR1233.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lillie S.H., Brown S.S.;
RL   Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 323-790 FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=94025982; PubMed=8212898;
RA   Schaaff-Gerstenschlaeger I., Bauer A., Boles E., Zimmermann F.K.;
RT   "Sequence and function analysis of a 4.3 kb fragment of Saccharomyces
RT   cerevisiae chromosome II including three open reading frames.";
RL   Yeast 9:915-921(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RA   Entian K.-D., Koetter P., Rose M., Becker J., Grey M., Li Z.,
RA   Niegemann E., Schenk-Groeninger R., Servos J., Wehner E.,
RA   Wolter R., Brendel M., Bauer J., Braun H., Dern K., Duesterhus S.,
RA   Gruenbein R., Hedges D., Kiesau P., Korol S., Krems B., Proft M.,
RA   Siegers K., Baur A., Boles E., Miosga T.,
RA   Schaaff-Gerstenschlaeger I., Zimmermann F.K.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: SUPPRESSOR OF THE MYO2 GENE.
CC   -!- SIMILARITY: SOME, TO S.POMBE SPAC4F10.13C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M90654; AAA35057.1; -.
DR   EMBL; X74437; CAA52452.1; -.
DR   EMBL; Z36041; CAA85133.1; -.
DR   PIR; S27458; S27458.
DR   SGD; S0000376; SMY2.
DR   InterPro; IPR003169; GYF.
DR   Pfam; PF02213; GYF; 1.
DR   SMART; SM00444; GYF; 1.
SQ   SEQUENCE   790 AA;  87098 MW;  99C0547FD1E68366 CRC64;


  Query Match             12.5%;  Score 75.5;  DB 1;  Length 790;
  Best Local Similarity   29.4%;  Pred. No. 31;
  Matches   35;  Conservative   13;  Mismatches   48;  Indels   23;  Gaps    5;

Qy       2 SQKQE----------EENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGS 51
             | |||          ::  :|   :   | |||     :| |||:|  |     ||
Db     402 SHKQEYKSTTMLEKGKKEKSESVAKALLDEQEKRNRELKRKEEARLSKK-----QKQKED 456

Qy      52 DFLMK-RLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSS 109
           | | | : || ||  :: :    |   | |:     |    | :   | :|||   ||
Db     457 DLLKKQKEQKEQKEKEALEAEKQKKSEKTKK-----DTQTQT-EGFKTSKDLPSLNSS 508


RESULT  24
NFM_RAT
ID   NFM_RAT        STANDARD;      PRT;   845 AA.
AC   P12839; Q63370;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neurofilament triplet M protein (160 kDa neurofilament protein)
DE   (Neurofilament medium polypeptide) (NF-M).
GN   NEF3 OR NEFM OR NFM.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87282618; PubMed=2441012;
RA   Napolitano E.W., Chin S.S.M., Colman D.R., Liem R.K.H.;
RT   "Complete amino acid sequence and in vitro expression of rat NF-M,
RT   the middle molecular weight neurofilament protein.";
```

```
RL   J. Neurosci. 7:2590-2599(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR;
RX   MEDLINE=92332596; PubMed=1321159;
RA   Kelly B.M., Gillespie C.S., Sherman D.L., Brophy P.J.;
RT   "Schwann cells of the myelin-forming phenotype express neurofilament
RT   protein NF-M.";
RL   J. Cell Biol. 118:397-410(1992).
RN   [3]
RP   PHOSPHORYLATION SITES, AND REVISION TO 500.
RX   MEDLINE=92165797; PubMed=1537832;
RA   Xu Z.-S., Liu W.-S., Willard M.B.;
RT   "Identification of six phosphorylation sites in the COOH-terminal
RT   tail region of the rat neurofilament protein M.";
RL   J. Biol. Chem. 267:4467-4471(1992).
RN   [4]
RP   CARBOHYDRATE-LINKAGE SITES.
RX   MEDLINE=93346421; PubMed=8344946;
RA   Dong D.L.-Y., Xu Z.-S., Chevrier M.R., Cotter R.J., Cleveland D.W.,
RA   Hart G.W.;
RT   "Glycosylation of mammalian neurofilaments. Localization of multiple
RT   O-linked N-acetylglucosamine moieties on neurofilament polypeptides
RT   L and M.";
RL   J. Biol. Chem. 268:16679-16687(1993).
CC   -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC       AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC   -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
CC       PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC       THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC       INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC       OF AXONAL CALIBER.
CC   -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC       OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC       LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC       COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M18628; AAA41696.1; -.
DR   EMBL; Z12152; CAA78136.1; -.
DR   GlycoSuiteDB; P12839; -.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW   Glycoprotein.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1    103       HEAD.
FT   DOMAIN      103    410       ROD.
FT   DOMAIN      411    844       TAIL.
FT   DOMAIN      103    134       COIL 1A.
FT   DOMAIN      135    147       LINKER 1.
FT   DOMAIN      148    246       COIL 1B.
FT   DOMAIN      247    263       LINKER 12.
FT   DOMAIN      264    285       COIL 2A.
FT   DOMAIN      286    289       LINKER 2.
FT   DOMAIN      290    410       COIL 2B.
FT   CARBOHYD     47     47       O-LINKED (GLCNAC).
FT                                /FTId=CAR_000130.
FT   CARBOHYD    430    430       O-LINKED (GLCNAC).
FT                                /FTId=CAR_000131.
FT   MOD_RES     502    502       PHOSPHORYLATION.
FT   MOD_RES     506    506       PHOSPHORYLATION.
FT   MOD_RES     536    536       PHOSPHORYLATION.
FT   MOD_RES     603    603       PHOSPHORYLATION.
FT   MOD_RES     608    608       PHOSPHORYLATION.
FT   MOD_RES     666    666       PHOSPHORYLATION.
FT   CONFLICT     17     17       MISSING (IN REF. 2).
FT   CONFLICT     21     21       R -> P (IN REF. 2).
FT   CONFLICT    204    204       V -> L (IN REF. 2).
FT   CONFLICT    500    500       MISSING (IN REF. 1).
SQ   SEQUENCE   845 AA;  95660 MW;  316C41655B11197D CRC64;


  Query Match           12.5%;  Score 75.5;  DB 1;  Length 845;
  Best Local Similarity  22.9%;  Pred. No. 34;
  Matches   25;  Conservative   23;  Mismatches   58;  Indels    3;  Gaps    2;

Qy       3 QKQEEENPAEETGEEKQDTQEK--EGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           |::| |  ||  ||| :  :||  || : | | : ::| : |   : |     : :   |
Db     557 QEEEGETEAEGEGEEAEAKEEKKTEGKVEEMAIKEEIKVEKPEKAKSPVPKSPVEEVKPK 616

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSS 109
             :     :   :  : |::   | :  |       |:|:| :| :
Db     617 PEAKAGKDEQKEEEKVEEKKEVAKESPKEEKVEKKE-EKPKDVPDKKKA 664


RESULT  25
NRT_DROME
ID   NRT_DROME      STANDARD;      PRT;   846 AA.
AC   P23654;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Neurotactin.
GN   NRT.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91301057; PubMed=2100266;
RA   Hortsch M., Patel N.P., Bieber A.J., Traquina Z.R., Goodman C.S.;
RT   "Drosophila neurotactin, a surface glycoprotein with homology to
RT   serine esterases, is dynamically expressed during embryogenesis.";
RL   Development 110:1327-1340(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R; TISSUE=Embryo;
RX   MEDLINE=91006059; PubMed=2120047;
RA   de la Escalera S., Bockamp E.O., Moya F., Piovant M., Jimenez F.;
RT   "Characterization and gene cloning of neurotactin, a Drosophila
RT   transmembrane protein related to cholinesterases.";
RL   EMBO J. 9:3593-3601(1990).
RN   [3]
RP   FUNCTION.
RX   MEDLINE=91006060; PubMed=2120048;
RA   Barthalay Y., Hipeau-Jacquotte R., de la Escalera S., Jimenez F.,
RA   Piovant M.;
RT   "Drosophila neurotactin mediates heterophilic cell adhesion.";
RL   EMBO J. 9:3603-3609(1990).
CC   -!- FUNCTION: MAY MEDIATE OR MODULATE CELL ADHESION BETWEEN EMBRYONIC
CC       CELLS DURING DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN MEMBRANES DURING EMBRYOGENESIS
CC       AND THEN ACCUMULATES IN CNS AND PNS.
CC   -!- PTM: PHOSPHORYLATED.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE TYPE-B
CC       CARBOXYLESTERASE/LIPASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X53837; CAA37831.1; -.
DR   EMBL; X54999; CAA38746.1; -.
DR   PIR; S12005; S12005.
DR   PIR; S13795; S13795.
DR   HSSP; P37967; 1QE3.
DR   FlyBase; FBgn0004108; Nrt.
DR   InterPro; IPR002018; Carboxylesterase_B.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   Pfam; PF00135; COesterase; 1.
DR   PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
KW   Cell adhesion; Glycoprotein; Transmembrane; Phosphorylation;
KW   Signal-anchor.
FT   DOMAIN        1    324       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    325    346       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN      347    846       EXTRACELLULAR (POTENTIAL).
FT   SIMILAR     351    846       WITH TYPE-B CARBOXYLESTERASE/LIPASES.
FT   MOD_RES      28     28       PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT   MOD_RES      75     75       PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT   MOD_RES     103    103       PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT   MOD_RES     169    169       PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT   MOD_RES     186    186       PHOSPHORYLATION (POTENTIAL).
FT   DISULFID    422    437       BY SIMILARITY.
FT   DISULFID    600    605       BY SIMILARITY.
FT   DISULFID    738    830       BY SIMILARITY.
FT   CARBOHYD    410    410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    428    428       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    636    636       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    691    691       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    720    720       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    250    250       G -> S (IN REF. 2).
FT   CONFLICT    555    555       S -> T (IN REF. 2).
FT   CONFLICT    567    567       F -> Y (IN REF. 2).
SQ   SEQUENCE   846 AA;  92745 MW;  F5EC54904D1D0657 CRC64;


  Query Match             12.5%;  Score 75.5;  DB 1;  Length 846;
  Best Local Similarity   24.4%;  Pred. No. 34;
  Matches   29;  Conservative   21;  Mismatches   50;  Indels   19;  Gaps    5;

Qy       1 MSQKQEEENPAEETGEEKQDT----QEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMK 56
           | : :|:| |  ||   :|:     :| : :| :: |:|| |   |    ||   :   |
Db       1 MGELEEKETPPTETTAAQQEALEEPKETDKML-DKKEDAKEKTPSPQT-SKPASPNAGKK 58

Qy      57 RLQKGQKYFDSGDYNMAKAK----------MKNKQLPSAGPDKNLVTGDHIPTPQDLP 104
             :|  |  |  :||:|            :|   | :  || :   :    |: :|
Db      59 SSPVAEKKID--DAELAKSKSGNGEEIIDIPAENGTKPDSADDKKISKEEREVKPKKIP 115


RESULT  26
FKB4_DROME
ID   FKB4_DROME     STANDARD;      PRT;   357 AA.
AC   P54397; Q9VF88;
DT   01-OCT-1996 (Rel. 34, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   39 kDa FK506-binding nuclear protein (Peptidyl-prolyl cis-trans
DE   isomerase) (PPiase) (EC 5.2.1.8).
GN   FK506-BP1 OR FKBP39 or CG6226.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE=95278752; PubMed=7538962;
RA   Theopold U., Dal Zotto L., Hultmark D.;
RT   "FKBP39, a Drosophila member of a family of proteins that bind the
RT   immunosuppressive drug FK506.";
RL   Gene 156:247-251(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS. FKBP506
CC       MAY FUNCTION IN A SIGNAL TRANSDUCTION CASCADE DURING EARLY
CC       DEVELOPMENT.
CC   -!- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC       PEPTIDE BONDS IN OLIGOPEPTIDES.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED, HIGHEST LEVELS IN
CC       OVARY.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED DURING ALL STAGES OF DEVELOPMENT
CC       WITH HIGHEST EXPRESSION IN EARLY EMBRYO.
CC   -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z46894; CAA86996.1; -.
DR   EMBL; AE003708; AAF55172.1; ALT_SEQ.
DR   HSSP; P27124; 1ROT.
DR   FlyBase; FBgn0013269; FK506-bp1.
DR   InterPro; IPR001179; FKBP_PPIase.
DR   Pfam; PF00254; FKBP; 1.
DR   PROSITE; PS00453; FKBP_PPIASE_1; FALSE_NEG.
DR   PROSITE; PS00454; FKBP_PPIASE_2; 1.
```

```
DR   PROSITE; PS50059; FKBP_PPIASE_3; 1.
KW   Isomerase; Rotamase; Nuclear protein.
FT   DOMAIN       89     99       ASP/GLU-RICH (HIGHLY ACIDIC).
FT   DOMAIN      119    183       ASP/GLU-RICH (HIGHLY ACIDIC).
FT   DOMAIN      186    247       LYS-RICH (BASIC).
FT   DOMAIN      269    357       PPIASE, FKBP-TYPE.
FT   CONFLICT    187    187       A -> R (IN REF. 1).
SQ   SEQUENCE   357 AA;  39343 MW;  EF0AB7831738BB30 CRC64;


  Query Match             12.4%;  Score 75;  DB 1;  Length 357;
  Best Local Similarity   32.4%;  Pred. No. 15;
  Matches   33;  Conservative   10;  Mismatches   35;  Indels   24;  Gaps    6;

Qy       6 EEENPAEETGEE-KQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKG--- 61
           :|:|  ||:||| :||: : |      ||| : |||   |   || |     : | |
Db     161 DEDN--EESGEEDEQDSDDSE------AEEEQPKAKVAKL--SPGASAKKSGKEQNGVAK 210

Qy      62 ---------QKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTG 94
                   :|          |||   || | |  |   :||
Db     211 KEEAKQQQKKKEKPEAKKEQPKAKEPAKQQP-ASKDPRTITG 251


RESULT  27
INCE_CHICK
ID   INCE_CHICK     STANDARD;      PRT;   877 AA.
AC   P53352;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Inner centromere protein (INCENP).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94012983; PubMed=8408220;
RA   Mackay A.M., Eckley D.M., Chue C., Earnshaw W.C.;
RT   "Molecular analysis of the INCENPs (inner centromere proteins):
RT   separate domains are required for association with microtubules
RT   during interphase and with the central spindle during anaphase.";
RL   J. Cell Biol. 123:373-385(1993).
CC   -!- FUNCTION: MAY ACT AS CYTOSKELETAL PROTEINS THAT ARE POTENTIALLY
CC       CAPABLE OF ALTERING THE MORPHOLOGY OF THE CELLULAR MICROTUBULE
CC       NETWORK DURING INTERPHASE.
CC   -!- SUBUNIT: HOMODIMER OR HETERODIMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR. IS RESTRICTED TO THE NUCLEUS IN
CC       INTERPHASE, REMAINS TIGHTLY BOUND TO THE CHROMOSOMES UNTIL EARLY
CC       METAPHASE, AND DURING LATE METAPHASE GETS CONCENTRATED IN LINEAR
CC       ARRAYS THAT TRANSECT THE METAPHASE PLATE BETWEEN THE CHROMOSOMES.
CC       AS ANAPHASE BEGINS TO MOVE TO THE SPINDLE MIDZONE WHERE IT IS
CC       INTIMATELY ASSOCIATED WITH THE BUNDLED MICROTUBULES. LATER IN
CC       ANAPHASE GETS CLOSELY ASSOCIATED WITH THE CELL CORTEX, AND BY
CC       TELOPHASE IS CONCENTRATED AT EACH SIDE OF THE MIDBODY IN THE
CC       INTERCELLULAR BRIDGE, WITH WHICH IT IS DISCARDED AFTER.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; INCENP I AND INCENP II (SHOWN
CC       HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z25419; CAA80906.1; -.
DR   EMBL; Z25420; CAA80907.1; -.
KW   Cell division; Microtubules; Coiled coil; Centromere; Mitosis;
KW   Cell cycle; Nuclear protein; Alternative splicing.
FT   DOMAIN      503    733       COILED COIL (POTENTIAL).
FT   VARSPLIC    716    753       MISSING (IN ISOFORM INCENP I).
FT   VARIANT     257    257       A -> T.
FT   VARIANT     471    471       E -> Q.
SQ   SEQUENCE   877 AA;  100940 MW;  AFA703149F555352 CRC64;


  Query Match             12.4%;  Score 75;  DB 1;  Length 877;
  Best Local Similarity   20.6%;  Pred. No. 39;
  Matches   27;  Conservative   28;  Mismatches   54;  Indels   22;  Gaps    3;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGIL-----------PERAEEAKLKAKYPSLGQKP-- 48
           :|:||:  | :  |::|  ||:|  |             : || :::|:     |:
Db     635 KKEEEQERARKIAEQRQAEQEREKQLAAEREQERKKEQERKKEEERIQAEKQREQQEKAA 694

Qy      49 -------GGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQ 101
                 : | | ::| :|  :     | : : : |:||     |::    |:   :
Db     695 RLQKEVLAAKEQLQKEMEKKEKE-EQLLAEMKRQEQEQKKLPEEQKAKDVAQTQHLENKE 753

Qy     102 DLPQRKSSLVT 112
           : |   |  :|
Db     754 NSPACNSYQMT 764


RESULT  28
ANK2_HUMAN
ID   ANK2_HUMAN     STANDARD;      PRT;  3924 AA.
AC   Q01484; Q01485;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).
GN   ANK2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC   TISSUE=Brain stem;
RX   MEDLINE=91302466; PubMed=1830053;
RA   Otto E., Kunimoto M., McLaughlin T., Bennett V.;
RT   "Isolation and characterization of cDNAs encoding human brain
RT   ankyrins reveal a family of alternatively spliced genes.";
RL   J. Cell Biol. 114:241-253(1991).
RN   [2]
RP   REVISIONS.
RA   Carpenter S.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain stem;
RX   MEDLINE=94075409; PubMed=8253844;
RA   Chan W., Kordeli E., Bennett V.;
RT   "440-kD ankyrinB: structure of the major developmentally regulated
RT   domain and selective localization in unmyelinated axons.";
RL   J. Cell Biol. 123:1463-1473(1993).
RN   [4]
RP   SEQUENCE OF 463-495 FROM N.A.
RX   MEDLINE=92009921; PubMed=1833308;
RA   Tse W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,
RA   Lux S.E., Ward D.C., Forget B.G.;
RT   "Isolation and chromosomal localization of a novel nonerythroid
RT   ankyrin gene.";
RL   Genomics 10:858-866(1991).
CC   -!- FUNCTION: Attach integral membrane proteins to cytoskeletal
CC       elements. Also bind to cytoskeletal proteins.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
CC       CELLS THROUGHOUT THE BRAIN.
CC   -!- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
CC       AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE
CC       AND FUNCTION (POTENTIAL).
```

```
CC   -!- SIMILARITY: CONTAINS 23 ANK REPEATS.
CC   -!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56957; CAA40278.1; -.
DR   EMBL; X56958; CAA40279.2; -.
DR   EMBL; Z26634; CAB42644.1; -.
DR   EMBL; M37123; AAA62828.1; -.
DR   PIR; S14533; S14533.
DR   PIR; A39643; A39643.
DR   PIR; B39643; B39643.
DR   PIR; S14569; S14569.
DR   HSSP; Q00421; 1AWC.
DR   MIM; 106410; -.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000488; Death.
DR   InterPro; IPR000906; ZU5.
DR   Pfam; PF00023; ank; 24.
DR   Pfam; PF00531; death; 1.
DR   Pfam; PF00791; ZU5; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 21.
DR   SMART; SM00005; DEATH; 1.
DR   SMART; SM00218; ZU5; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 20.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50017; DEATH_DOMAIN; 1.
KW   Cytoskeleton; Alternative splicing; Repeat; ANK repeat;
KW   Phosphorylation.
FT   REPEAT       63     92       ANK 1.
FT   REPEAT       96    125       ANK 2.
FT   REPEAT      129    158       ANK 3.
FT   REPEAT      162    191       ANK 4.
FT   REPEAT      193    220       ANK 5.
FT   REPEAT      232    261       ANK 6.
FT   REPEAT      265    294       ANK 7.
FT   REPEAT      298    327       ANK 8.
FT   REPEAT      331    360       ANK 9.
FT   REPEAT      364    393       ANK 10.
FT   REPEAT      397    426       ANK 11.
FT   REPEAT      430    459       ANK 12.
FT   REPEAT      463    492       ANK 13.
FT   REPEAT      496    525       ANK 14.
FT   REPEAT      529    558       ANK 15.
FT   REPEAT      562    591       ANK 16.
FT   REPEAT      595    624       ANK 17.
FT   REPEAT      628    657       ANK 18.
FT   REPEAT      661    690       ANK 19.
FT   REPEAT      694    723       ANK 20.
FT   REPEAT      727    756       ANK 21.
FT   REPEAT      760    789       ANK 22.
FT   REPEAT      793    822       ANK 23.
FT   DOMAIN     1773   1950       REPEAT-RICH REGION.
FT   REPEAT     1773   1784       REPEAT A.
FT   REPEAT     1785   1796       REPEAT A.
FT   REPEAT     1797   1808       REPEAT A.
FT   REPEAT     1809   1820       REPEAT A.
FT   REPEAT     1821   1832       REPEAT A.
FT   REPEAT     1833   1844       REPEAT A.
FT   REPEAT     1845   1856       REPEAT A.
FT   REPEAT     1857   1867       REPEAT A (APPROXIMATE).
FT   REPEAT     1868   1879       REPEAT A.
FT   REPEAT     1880   1891       REPEAT A.
FT   REPEAT     1892   1902       REPEAT A (APPROXIMATE).
FT   REPEAT     1903   1914       REPEAT A.
FT   REPEAT     1915   1926       REPEAT A.
FT   REPEAT     1927   1938       REPEAT A.
FT   REPEAT     1939   1950       REPEAT A.
FT   DOMAIN     3536   3620       DEATH.
FT   VARSPLIC   1039   1039       Q -> QFLGKLHLPTAPPPLNEGESLVSRILQLGPPGTK
FT                                (IN ISOFORM 2).
FT   VARSPLIC   1444   3528       MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT   CONFLICT    475    476       GQ -> PE (IN REF. 4).
FT   CONFLICT    971    971       I -> S (IN REF. 1).
FT   CONFLICT   3581   3582       QY -> HA (IN REF. 1).
FT   CONFLICT   3586   3586       I -> Y (IN REF. 1).
SQ   SEQUENCE   3924 AA;  430337 MW;  52AC496C428E29D2 CRC64;


  Query Match             12.4%;  Score 75;  DB 1;  Length 3924;
  Best Local Similarity   30.5%;  Pred. No. 1.9e+02;
  Matches   39;  Conservative   21;  Mismatches   48;  Indels   20;  Gaps    8;

Qy       3 QKQEEEN---PAEETGEEKQDTQEKEGILPERAEE--AKLKAKYPSLGQKPGGS-----D 52
           |||:||     ||:   :|  ::|  |  |  | |    :||  | : : | ||
Db    1679 QKQKEEGLQASAEKAELKKGSSEESLGEDPGLAPEPLPTVKATSPLIEETPIGSIKDKVK 1738

Qy      53 FLMKRL---QKGQKYFD---SGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIP-TPQDLPQ 105
            | ||:   |||:         |  :: | |  ::  |:| |  :| :  | | :|    :
Db    1739 ALQKRVEDEQKGRSKLPIRVKGKEDVPK-KTTHRPHPAASP--SLKSERHAPGSPSPKTE 1795

Qy     106 RKSSLVTS 113
           | |:| :|
Db    1796 RHSTLSSS 1803


RESULT  29
NUKS_RAT
ID   NUKS_RAT       STANDARD;      PRT;   243 AA.
AC   Q9EPJ0;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nuclear ubiquitous casein and cyclin-dependent kinases substrate.
GN   NUCKS.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RA   Oestvold A.C., Wanvik B., Mathiesen S., Norum J.H., Nordeide H.,
RA   Olstad O.K.;
RT   "Molecular cloning of a novel mammalian nuclear phosphoprotein serving
RT   as substrate for CDK1 in vivo.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- PTM: PHOSPHORYLATED BY CDK1 AND CASEIN KINASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ237669; CAC20862.1; -.
KW   Phosphorylation; Nuclear protein.
FT   DOMAIN      140    145       POLY-ASP.
FT   DOMAIN      150    206       LYS-RICH.
SQ   SEQUENCE   243 AA;  27140 MW;  98B5BB2A7F2133C8 CRC64;


  Query Match             12.3%;  Score 74.5;  DB 1;  Length 243;
  Best Local Similarity   22.3%;  Pred. No. 11;
  Matches   31;  Conservative   26;  Mismatches   35;  Indels   47;  Gaps    7;
```

```
Qy       4 KQEEENPAEETGEEKQD--------------TQEKEGIL--------PERAEEAKLKAK 40
           |::: : ||:: :|| |               ::::| :|        ||  :||  :
Db      69 KKDDSHSAEDSEDEKDDHKSVRQQRQAASKAASKQREMLLEDVGSEEEPEEDDEAPFQ-- 126

Qy      41 YPSLGQKPGGS--DFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIP 98
                :|  ||  |||:         |  ||  :| | | | :  : |::
Db     127 -----EKDSGSDEDFLMED-------DDDSDYGSSKKKNK-KMVKKSKPERK-------E 166

Qy      99 TPQDLPQRKSSLVTSKLAG 117
                |: |:::  | : |
Db     167 KKMPKPRLKATVTPSPVKG 185


RESULT  30
DD21_MOUSE
ID   DD21_MOUSE     STANDARD;      PRT;   851 AA.
AC   Q9JIK5; Q9WV45;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Nucleolar RNA helicase II (Nucleolar RNA helicase Gu) (RH II/Gu)
DE   (DEAD-box protein 21).
GN   DDX21.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ;
RX   MEDLINE=20318620; PubMed=10860663;
RA   Valdez B.C., Wang W.;
RT   "Mouse RNA helicase II/Gu: cDNA and genomic sequences, chromosomal
RT   localization, and regulation of expression.";
RL   Genomics 66:184-194(2000).
RN   [2]
RP   REVISIONS OF N-TERMINUS.
RA   Valdez B.C.;
RL   Submitted (APR-2001) to the SWISS-PROT data bank.
CC   -!- FUNCTION: CAN UNWIND DOUBLE-STRANDED RNA (HELICASE) AND CAN FOLD
CC       OR INTRODUCE A SECONDARY STRUCTURE TO A SINGLE-STRANDED RNA
CC       (FOLDASE).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LIVER AND TESTIS.
CC       EXPRESSED AT LOWER LEVEL IN BRAIN, LUNGS, AND SKELETAL MUSCLE.
CC   -!- DOMAIN: THE TWO ENZYMATIC ACTIVITIES RESIDE IN TWO SEPARATE
CC       DOMAINS, THE HELICASE IN THE N-TERMINUS AND THE FOLDASE IN THE C-
CC       TERMINUS.
CC   -!- DOMAIN: THE 3 X 5 AA REPEATS SEEM TO BE CRITICAL FOR THE RNA
CC       FOLDING ACTIVITY.
CC   -!- DOMAIN: THE 3 X 37 AA TANDEM REPEATS MAY BE IMPORTANT FOR NUCLEAR
CC       LOCALIZATION AND/OR INTERACTION WITH NUCLEIC ACIDS OR OTHER
CC       PROTEINS.
CC   -!- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF220365; AAF61690.1; -.
DR   EMBL; AF159131; AAD43959.3; -.
DR   MGD; MGI:1860494; Ddx21.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR000629; DEAD_ATP_helicase.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00039; DEAD_ATP_HELICASE; FALSE_NEG.
KW   Helicase; RNA-binding; ATP-binding; Nuclear protein; Antigen; Repeat.
FT   NP_BIND     302    309       ATP (POTENTIAL).
FT   SITE        411    414       DEVD BOX.
FT   DOMAIN       95     98       POLY-LYS.
FT   DOMAIN      117    227       3 X 37 AA TANDEM REPEATS.
FT   REPEAT      117    153       1-1.
FT   REPEAT      154    190       1-2.
FT   REPEAT      191    227       1-3.
FT   DOMAIN      807    833       3 X 5 AA REPEATS.
FT   REPEAT      807    811       2-1.
FT   REPEAT      817    823       2-2.
FT   REPEAT      829    833       2-3.
FT   CONFLICT      1     46       MISSING (IN REF. 1; AAF61690).
FT   CONFLICT    790    790       Q -> P (IN REF. 1; AAF61690).
SQ   SEQUENCE   851 AA;  93582 MW;  71D7286F4CA1DDD7 CRC64;


  Query Match            12.3%;  Score 74.5;  DB 1;  Length 851;
  Best Local Similarity   25.2%;  Pred. No. 41;
  Matches   30;  Conservative   22;  Mismatches   32;  Indels   35;  Gaps    6;

Qy       4 KQEEENPAEETG------------EEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGG 50
           |::|| | ::|              |:: |:: || |  | :||   :|  |    ||
Db      77 KKQEEEPQDDTASTSKTSKKKKEPLEKQADSETKEIITEEPSEE---EADMP----KP-- 127

Qy      51 SDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSS 109
                |:::||::   :||      |:||      |  :       || |: : : | :
Db     128 -----KKMKKGKE--ANGDAGEKSPKLKN------GLSQPSEEEADIPKPKKMKKGKEA 173


RESULT  31
CHD2_HUMAN
ID   CHD2_HUMAN     STANDARD;      PRT;  1739 AA.
AC   O14647;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Chromodomain-helicase-DNA-binding protein 2 (CHD-2).
GN   CHD2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97470991; PubMed=9326634;
RA   Woodage T., Basrai M.A., Baxevanis A.D., Hieter P., Collins F.S.;
RT   "Characterization of the CHD family of proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997).
CC   -!- FUNCTION: SEQUENCE-SELECTIVE DNA-BINDING PROTEIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 'CHROMO' DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF006514; AAB87382.1; -.
DR   HSSP; P23197; 1AP0.
DR   MIM; 602119; -.
DR   InterPro; IPR000953; Chromo.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00385; chromo; 2.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
```

```
DR   SMART; SM00298; CHROMO; 2.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00598; CHROMO_1; 2.
DR   PROSITE; PS50013; CHROMO_2; 2.
KW   DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat.
FT   DOMAIN        1     74       SER-RICH.
FT   DOMAIN      121    139       SER-RICH.
FT   DOMAIN      261    353       CHROMO 1.
FT   DOMAIN      378    456       CHROMO 2.
FT   NP_BIND     509    516       ATP (POTENTIAL).
FT   SITE        617    620       DEAH BOX.
SQ   SEQUENCE   1739 AA;  200560 MW;  D62089C25EBAA40F CRC64;


  Query Match             12.3%;  Score 74.5;  DB 1;  Length 1739;
  Best Local Similarity   29.3%;  Pred. No. 89;
  Matches   36;  Conservative   17;  Mismatches   53;  Indels   17;  Gaps    7;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGI--LPERAEEAKLKAKYPSLGQKPGGSDFLMKRL 58
           :|  :  :||:|| || | |  ||  :    :: |  : | |  |  :   |     |
Db    1363 LSSPRHSDNPSEE-GEVKDDGLEKSPMKKKQKKKENKENKEKQMSSRKDKEGDKERKKSK 1421

Qy      59 QKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTP------QDLPQRKSSLVT 112
           | :|   |||   ||:  |:|:  | ||    :|    | |       || |   |:
Db    1422 DKKEKP-KSGD---AKSSSKSKR--SQGPVH--ITAGSEPVPIGEDEDDDLDQETFSICK 1473

Qy     113 SKL 115
           ::
Db    1474 ERM 1476


RESULT  32
PRKD_HUMAN
ID   PRKD_HUMAN     STANDARD;      PRT;  4128 AA.
AC   P78527; Q14175; P78528; Q13337; Q13327;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA-dependent protein kinase catalytic subunit (EC 2.7.1.37) (DNA-
DE   PKcs) (DNPK1).
GN   PRKDC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cervical carcinoma;
RX   MEDLINE=95401275; PubMed=7671312;
RA   Hartley K.O., Gell D., Smith G.C.M., Zhang H., Divecha N.,
RA   Connelly M.A., Admon A., Lees-Miller S.P., Anderson C.W.,
RA   Jackson S.P.;
RT   "DNA-dependent protein kinase catalytic subunit: a relative of
RT   phosphatidylinositol 3-kinase and the ataxia telangiectasia gene
RT   product.";
RL   Cell 82:849-856(1995).
RN   [2]
RP   REVISIONS.
RA   Gell D., Anderson C.W.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-1689 FROM N.A.
RA   Anderson C.W., Freimuth P.I., Dunn J.J.;
RT   "Partial sequence of a BAC insert containing exons 8 through 38 of the
RT   human gene encoding DNA-PKcs.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1789-2203 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=95365397; PubMed=7638222;
RA   Sipley J.D., Menninger J.C., Hartley K.O., Ward D.C., Jackson S.P.,
RA   Anderson C.W.;
RT   "Gene for the catalytic subunit of the human DNA-activated protein
RT   kinase maps to the site of the XRCC7 gene on chromosome 8.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:7515-7519(1995).
RN   [5]
RP   SEQUENCE OF 3199-4128 FROM N.A.
RC   TISSUE=Fetal lung;
RX   MEDLINE=96062265; PubMed=7594449;
RA   Poltoratsky V.P., Shi X., York J.D., Lieber M.R., Carter T.H.;
RT   "Human DNA-activated protein kinase (DNA-PK) is homologous to
RT   phosphatidylinositol kinases.";
RL   J. Immunol. 155:4529-4533(1995).
RN   [6]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=97074683; PubMed=8917110;
RA   Connelly M.A., Zhang H., Kieleczawa J., Anderson C.W.;
RT   "Alternate splice-site utilization in the gene for the catalytic
RT   subunit of the DNA-activated protein kinase, DNA-PKcs.";
RL   Gene 175:271-273(1996).
RN   [7]
RP   CHARACTERIZATION.
RX   MEDLINE=98040126; PubMed=9372844;
RA   Wu X., Lieber M.R.;
RT   "Interaction between DNA-dependent protein kinase and a novel protein,
RT   KIP.";
RL   Mutat. Res. 385:13-20(1997).
CC   -!- FUNCTION: SER/THR KINASE INVOLVED IN DNA DOUBLE-STRANDED BREAK
CC       REPAIR, V(D)J RECOMBINATION AND MODULATION OF TRANSCRIPTION. MUST
CC       BE BOUND TO DNA TO EXPRESS ITS CATALYTIC PROPERTIES.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SUBUNIT: FORMS A COMPLEX WITH THE HETERODIMER KU70/KU80
CC       AUTOANTIGEN. INTERACTS WITH DNA-PKCS INTERACTING PROTEIN (KIP)
CC       WITH THE REGION UPSTREAM THE KINASE DOMAIN.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U47077; AAB39925.3; -.
DR   EMBL; U34994; AAC50210.2; -.
DR   EMBL; U63630; AAC52019.2; -.
DR   EMBL; U35835; AAA79184.1; -.
DR   EMBL; L27425; AAA79244.1; -.
DR   MIM; 600899; -.
DR   InterPro; IPR003151; FAT.
DR   InterPro; IPR003152; FATC.
DR   InterPro; IPR000403; PI3_PI4_kinase.
DR   Pfam; PF02259; FAT; 1.
DR   Pfam; PF02260; FATC; 1.
DR   Pfam; PF00454; PI3_PI4_kinase; 1.
DR   SMART; SM00146; PI3Kc; 1.
DR   PROSITE; PS00915; PI3_4_KINASE_1; 1.
DR   PROSITE; PS00916; PI3_4_KINASE_2; FALSE_NEG.
DR   PROSITE; PS50290; PI3_4_KINASE_3; 1.
KW   Transferase; Serine/threonine-protein kinase; DNA repair;
KW   Nuclear protein; Alternative splicing.
FT   DOMAIN     2436   3212       KIP BINDING.
FT   DOMAIN     3748   4128       PI3K/PI4K.
FT   VARSPLIC   3799   3829       MISSING (IN ISOFORM 2).
FT   CONFLICT    405    405       Y -> D (IN REF. 3).
FT   CONFLICT   1008   1008       S -> A (IN REF. 3).
FT   CONFLICT   3434   3434       T -> I (IN REF. 5).
FT   CONFLICT   3660   3660       N -> T (IN REF. 5).
FT   CONFLICT   3817   3817       L -> W (IN REF. 5).
FT   CONFLICT   3862   3862       A -> P (IN REF. 5).
SQ   SEQUENCE   4128 AA;  469136 MW;  23E1861CE68606F4 CRC64;
```

```
  Query Match             12.3%;  Score 74.5;  DB 1;  Length 4128;
  Best Local Similarity   25.7%;  Pred. No. 2.2e+02;
  Matches   27;  Conservative   14;  Mismatches   37;  Indels   27;  Gaps    5;

Qy      20 DTQEKEGILPERAEEAKLKAKYPSLGQ---KPGGSDFLMKRLQKGQKYFD---------- 66
Db    2608 ETQASQGTLQTRTQEGSLSARWPVAGQIRATQQQHDFTLTQTADGRSSFDWLTGSSTDPL 2667

Qy      67 ------SGD----YNMAKAKMKNKQLPSAGPD---KNL-VTGDHI 97
Db    2668 VDHTSPSSDSLLFAHKRSERLQRAPLKSVGPDFGKKRLGLPGDEV 2712


RESULT  33
RDGB_DROME
ID   RDGB_DROME     STANDARD;      PRT;  1054 AA.
AC   P43125;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Retinal degeneration B protein (Probable calcium transporter RDGB).
GN   RDGB.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R; TISSUE=Head;
RX   MEDLINE=91231170; PubMed=1903119;
RA   Vihtelic T.S., Hyde D.R., O'Tousa J.E.;
RT   "Isolation and characterization of the Drosophila retinal
RT   degeneration B (rdgB) gene.";
RL   Genetics 127:761-768(1991).
CC   -!- FUNCTION: RDGB MUTANTS UNDERGO RAPID LIGHT-INDUCED RETINAL
CC       DEGENERATION. MAY CONTROL PHOSPHATIDYLINOSITOL CONCENTRATION IN
CC       TRANSPORT VESICLES FROM THE SUBRHABDOMERIC CISTERNAE (SRC) TO THE
CC       RHABDOMERE. MAY FUNCTION AS A CALCIUM TRANSPORTER.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN ADULT HEADS, NOT DETECTED IN
CC       BODIES.
CC   -!- SIMILARITY: THE N-TERMINAL IS SIMILAR TO MAMMALIAN
CC       PHOSPHATIDYLINOSITOL (PTDINS) TRANSFER PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57978; CAA41044.1; -.
DR   PIR; A61221; A61221.
DR   FlyBase; FBgn0003218; rdgB.
DR   InterPro; IPR004177; DDHD.
DR   InterPro; IPR001666; PI_transfer.
DR   Pfam; PF02862; DDHD; 1.
DR   Pfam; PF02121; IP_trans; 1.
DR   PRINTS; PR00391; PITRANSFER.
KW   Transmembrane; Vision.
FT   DOMAIN      321    332       ASP/GLU-RICH (HIGHLY ACIDIC).
FT   TRANSMEM    501    519       POTENTIAL.
FT   TRANSMEM    734    750       POTENTIAL.
FT   CARBOHYD    194    194       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    414    414       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    612    612       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    658    658       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    852    852       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    928    928       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1054 AA;  116560 MW;  87F26EB871003CA8 CRC64;

  Query Match             12.1%;  Score 73.5;  DB 1;  Length 1054;
  Best Local Similarity   23.8%;  Pred. No. 63;
  Matches   34;  Conservative   18;  Mismatches   50;  Indels   41;  Gaps    6;

Qy       3 QKQEEENPAEETGEEKQDTQEKE-------------GILPERAEEAKL----KAKYPSL 44
Db     311 QPPSAEASSDEEGEEEEDDDEDENDAIGTGVDLSANQGGSAQRSRSQSIQMAQKGKFGSK 370

Qy      45 G--QKPGGS---------DFLMKRLQ------KGQKYFDSGDYNMAKAKMKNKQLPSAGP 87
Db     371 GALHSPVGSAHSFDLQVANWRMERLEVDSKSNSDEEFFDCLDTNETNSLAKWSSLELLG- 429

Qy      88 DKNLVTGDHIPTPQDLPQRKSSL 110
Db     430 -----EGDDSPPPHGGPSSAASV 447


RESULT  34
HEX_ADE01
ID   HEX_ADE01      STANDARD;      PRT;   467 AA.
AC   Q04965;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hexon protein (Late protein 2) (Fragment).
GN   PII.
OS   Human adenovirus type 1.
OC   Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX   NCBI_TaxID=10533;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93289026; PubMed=8511397;
RA   Pring-Akerblom P., Adrian T.;
RT   "The hexon genes of adenoviruses of subgenus C: comparison of the
RT   variable regions.";
RL   Res. Virol. 144:117-127(1993).
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF THE STRUCTURAL PROTEINS IN THE
CC       VIRAL COAT AND IS SYNTHESIZED DURING LATE INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67709; CAA47946.1; -.
DR   HSSP; P03277; 1DHX.
DR   InterPro; IPR000736; Adeno_hexon.
DR   Pfam; PF01065; Adeno_hexon; 1.
DR   ProDom; PD002815; Adeno_hexon; 1.
KW   Coat protein; Hexon protein; Late protein.
FT   NON_TER       1      1
FT   DOMAIN       15     45       GLU-RICH (ACIDIC).
FT   NON_TER     467    467
SQ   SEQUENCE   467 AA;  52060 MW;  DC72E5BBFA6FB825 CRC64;


  Query Match             12.1%;  Score 73;  DB 1;  Length 467;
  Best Local Similarity   26.9%;  Pred. No. 29;
  Matches   29;  Conservative   16;  Mismatches   29;  Indels   34;  Gaps    5;

Qy       6 EEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
Db      17 EQEEPTQEMAGELEDEEEAE------EEEAEEEAEAPQAGQK----------VKKTHVY- 59

Qy      66 DSGDYNMAKAKMKNKQLPSAG--------PDKNLVTGDHIPTPQDLPQ 105
Db      60 -------AQAPLAGEKITANGLQIVSDTQTEGNPVFAD--PTYQPEPQ 98
```

```
RESULT  35
HAP1_RAT
ID   HAP1_RAT       STANDARD;      PRT;   629 AA.
AC   P54256;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Huntingtin associated protein 1 (HAP1).
GN   HAP1-A AND HAP1-B.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=96077112; PubMed=7477378;
RA   Li X.-J., Li S.-H., Sharp A.H., Nucifora F.C. Jr., Schilling G.,
RA   Lanahan A., Worley P., Snyder S.H., Ross C.A.;
RT   "A huntingtin-associated protein enriched in brain with implications
RT   for pathology.";
RL   Nature 378:398-402(1995).
CC   -!- FUNCTION: ASSOCIATES SPECIFICALLY WITH HUNTINGTIN. THIS BINDING IS
CC       ENHANCED BY AN EXPANDED POLYGLUTAMINE REPEAT.
CC   -!- ALTERNATIVE PRODUCTS: ALTERNATIVE SPLICING PRODUCES A SHORT FORM
CC       (A) AND A LONG FORM (B). THE LATEST IS SHOWN HERE.
CC   -!- TISSUE SPECIFICITY: IN THE BRAIN, SPECIALLY IN THE OLFACTORY BULB
CC       AND IN THE BRAIN STEM. NO DETECTABLE EXPRESSION IN PERIPHERAL
CC       TISSUES SUCH AS LUNG, TESTIS, SPLEEN, AND SMALL INTESTINE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U38370; AAC52326.1; -.
DR   EMBL; U38373; AAC52327.1; -.
KW   Alternative splicing.
FT   DOMAIN      220    237       POLY-GLU.
FT   DOMAIN      468    471       POLY-PRO.
FT   VARSPLIC    579    599       DSPAPQQQTNMGGGIVEQQPI -> GECSRRGHPPASGTSY
FT                                RSSTL (IN ISOFORM A).
FT   VARSPLIC    600    629       MISSING (IN ISOFORM A).
SQ   SEQUENCE   629 AA;  70213 MW;  C23C5CB57520B574 CRC64;


  Query Match             12.1%;  Score 73;  DB 1;  Length 629;
  Best Local Similarity   25.8%;  Pred. No. 40;
  Matches   24;  Conservative   18;  Mismatches   43;  Indels    8;  Gaps    3;

Qy       7 EENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL---GQKPGGSD--FLMKRLQKG 61
           |: ||||| :  ::|:  | : ||  |  ::      :|   |  |   |  :::::|
Db     503 EDGPAEETEQASEETEAWEEVEPEVDEATRMNVVVSALEASGLGPSHLDMKYVLQQLSNW 562

Qy      62 QKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTG 94
           |   |:      | |:  |  |:     |:  |
Db     563 Q---DAHSKRQQKQKVVPKDSPAPQQQTNMGGG 592


RESULT  36
TRDN_RABIT
ID   TRDN_RABIT     STANDARD;      PRT;   705 AA.
AC   Q28820; Q28636; Q28637; Q28643;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Triadin.
GN   TRDN.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (SKELETAL ISOFORMS).
RC   STRAIN=NEW ZEALAND WHITE; TISSUE=Skeletal muscle;
RX   MEDLINE=93286104; PubMed=7685347;
RA   Knudson C.M., Stang K.K., Moomaw C.R., Slaughter C.A., Campbell K.P.;
RT   "Primary structure and topological analysis of a skeletal muscle-
RT   specific junctional sarcoplasmic reticulum glycoprotein (triadin).";
RL   J. Biol. Chem. 268:12646-12654(1993).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (SKELETAL ISOFORMS).
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=94298946; PubMed=8026576;
RA   Peng M., Fan H., Kirley T.L., Caswell A.H., Schwartz A.;
RT   "Structural diversity of triadin in skeletal muscle and evidence of
RT   its existence in heart.";
RL   FEBS Lett. 348:17-20(1994).
RN   [3]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (CARDIAC ISOFORMS).
RC   TISSUE=Heart muscle;
RX   MEDLINE=96132942; PubMed=8550602;
RA   Guo W., Jorgensen A.O., Jones L.R., Campbell K.P.;
RT   "Biochemical characterization and molecular cloning of cardiac
RT   triadin.";
RL   J. Biol. Chem. 271:458-465(1996).
CC   -!- FUNCTION: MAY BE INVOLVED IN ANCHORING CALSEQUESTRIN TO THE
CC       JUNCTIONAL SARCOPLASMIC RETICULUM AND ALLOWING ITS FUNCTIONAL
CC       COUPLING WITH THE RYANODINE RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. SARCOPLASMIC
CC       RETICULUM.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 6 ISOFORMS; CARDIAC ISOFORM 1/CT1,
CC       CARDIAC ISOFORM 2/CT2, CARDIAC ISOFORM 3/CT3, SKELETAL ISOFORM
CC       1/ST1 (SHOWN HERE), SKELETAL ISOFORM 2/ST2 AND SKELETAL ISOFORM
CC       3/ST3; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: SKELETAL AND CARDIAC MUSCLE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U31540; AAC48496.1; -.
DR   EMBL; L10065; AAA31488.1; -.
DR   EMBL; U31555; AAC48497.1; -.
DR   EMBL; U34201; AAC48498.1; -.
KW   Transmembrane; Sarcoplasmic reticulum; Glycoprotein;
KW   Alternative splicing.
FT   INIT_MET      0      0
FT   DOMAIN        1     46       CYTOPLASMIC.
FT   TRANSMEM     47     67       POTENTIAL.
FT   DOMAIN       68    705       LUMENAL.
FT   CARBOHYD     74     74       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    624    624       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    264    285       DQYAFCRYMIDIFVHGDLKPGQ -> GGKQSEEAAGCFKRT
FT                                LGKKQMQ (IN CARDIAC ISOFORM 1).
FT   VARSPLIC    264    307       DQYAFCRYMIDIFVHGDLKPGQSPAIPPPSPTEQASRPTPA
FT                                LPT -> ECIFLSAATPQGIPNRQQLNDIHHCFLKTKKGGN
FT                                GQHAFCLKGC (IN CARDIAC ISOFORM 2).
FT   VARSPLIC    286    705       MISSING (IN CARDIAC ISOFORM 1).
FT   VARSPLIC    308    705       MISSING (IN CARDIAC ISOFORM 2).
FT   VARSPLIC    415    415       A -> E (IN SKELETAL ISOFORM 2).
FT   VARSPLIC    416    424       MISSING (IN SKELETAL ISOFORM 2).
FT   VARSPLIC    584    584       D -> E (IN SKELETAL ISOFORM 2 AND
FT                                SKELETAL ISOFORM 3).
FT   VARSPLIC    585    591       MISSING (IN SKELETAL ISOFORM 2 AND
FT                                SKELETAL ISOFORM 3).
FT   VARSPLIC    645    705       SKKAKEEAEEVSSTKKQKSPISFFQCVYLDGYNGYGFQFPV
```

```
FT                              TPAQYPGESSGKPNSPGPKQ -> LLATVGIWGMNQWMEDL
FT                              SVTLPSK (IN CARDIAC ISOFORM 3).
SQ   SEQUENCE   705 AA;  79003 MW;  13AF1D84475A1361 CRC64;


  Query Match             12.1%;  Score 73;  DB 1;  Length 705;
  Best Local Similarity   22.7%;  Pred. No. 45;
  Matches   32;  Conservative   20;  Mismatches   39;  Indels   50;  Gaps    5;

Qy       2 SQKQEEENPA---------------EETGEEKQDTQEKEGILPERAEEAKL----KAKY 41
           :: ||:|ǁ:                 |:|::| :  |||  :| | |  :|    ||:
Db     571 AEHQEKEPPSIKTDKPKSTSKGMPEVTESGKKKIEKSEKEIKVPARRESHQLQNVTKAEK 630

Qy      42 PSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVT-------- 93
           |: | | |  |    :               ||| : ::: |    |: ::
Db     631 PARGSKEGFEDVPASK---------------KAKEEAEEVSSTKKQKSPISFFQCVYLD 674

Qy      94 -----GDHIP-TPQDLPQRKS 108
                |   | ||   |   |
Db     675 GYNGYGFQFPVTPAQYPGESS 695


RESULT  37
YMT9_YEAST
ID   YMT9_YEAST     STANDARD;      PRT;   807 AA.
AC   Q04660;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 91.7 kDa Trp-Asp repeats containing protein IN
DE   NUP116-FAR3 intergenic region.
GN   YMR049C OR YM9796.02C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
CC   -!- SIMILARITY: STRONG, TO HUMAN KIAA0124.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49703; CAA89759.1; -.
DR   SGD; S0004652; YMR049C.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 5.
DR   PRINTS; PR00320; GPROTEINBRPT.
DR   SMART; SM00320; WD40; 5.
DR   PROSITE; PS00678; WD_REPEATS_1; 1.
DR   PROSITE; PS50082; WD_REPEATS_2; 2.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 2.
KW   Hypothetical protein; Repeat; WD repeat.
FT   REPEAT      435    474       WD 1.
FT   REPEAT      483    523       WD 2.
FT   REPEAT      592    634       WD 3.
FT   REPEAT      678    717       WD 4.
FT   REPEAT      721    760       WD 5.
FT   REPEAT      776    806       WD 6.
SQ   SEQUENCE   807 AA;  91704 MW;  6542DFE4670341CA CRC64;


  Query Match             12.1%;  Score 73;  DB 1;  Length 807;
  Best Local Similarity   29.5%;  Pred. No. 52;
  Matches   33;  Conservative   20;  Mismatches   39;  Indels   20;  Gaps    7;

Qy       5 QEEENPAEE---TGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKG 61
           :|  || ||   : |||:  :  |    || |   :  || :| : ||  : : :| ::
Db     335 EESYNPPEEYLLSPEEKEAWENTE--YSER-ERNFIPQKYSALRKVPGYGESIRERFER- 390

Qy      62 QKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDL---PQRKSSL 110
                | |  :|   |||       | | :  : :|:|:||   | | |::
Db     391 -----SLDLYLAPRVRKNK----LNIDPNSLIPE-LPSPKDLRPFPIRCSTI 432


RESULT  38
DNM1_HUMAN
ID   DNM1_HUMAN     STANDARD;      PRT;  1616 AA.
AC   P26358; Q9UHG5; Q9ULA2; Q9UMZ6;
DT   01-MAY-1992 (Rel. 22, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   DNA (cytosine-5)-methyltransferase 1 (EC 2.1.1.37) (DNA
DE   methyltransferase HsaI) (DNA MTase HsaI) (MCMT) (M.HsaI).
GN   DNMT1 OR DNMT OR AIM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=92279022; PubMed=1594447;
RA   Yen R.W.C., Vertino P.M., Nelkin B.D., Yu J.J., Deiry W.E.,
RA   Cumaraswamy A., Lennon G.G., Trask B.J., Celano P., Baylin S.B.;
RT   "Isolation and characterization of the cDNA encoding human DNA
RT   methyltransferase.";
RL   Nucleic Acids Res. 20:2287-2291(1992).
RN   [2]
RP   REVISIONS TO N-TERMINUS.
RX   MEDLINE=97094871; PubMed=8940105;
RA   Yoder J.A., Yen R.-W.C., Vertino P.M., Bestor T.H., Baylin S.B.;
RT   "New 5' regions of the murine and human genes for DNA (cytosine-5)-
RT   methyltransferase.";
RL   J. Biol. Chem. 271:31092-31097(1996).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RC   TISSUE=Prostatic carcinoma;
RA   Li L.C., Au H., Chui R., Dahiya R.;
RT   "Human DNA methyltransferase (DNMT1) is alternatively spliced.";
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 122-1087 FROM N.A.
RA   Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA   Burkhart-Schultz K., Gordon L., Dias J., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA   Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA   Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,
RA   Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
RA   Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA   Olsen A.S., Carrano A.V.;
RT   "Sequence analysis of a 6 Mb region in 19p13.2 between D19S391 and
RT   D19S179.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   PARTIAL SEQUENCE FROM N.A. (ISOFORM 2).
RX   MEDLINE=99380591; PubMed=10449766;
RA   Hsu D.-W., Lin M.-J., Lee T.-L., Wen S.-C., Chen X., Shen C.-J.;
RT   "Two major forms of DNA (cytosine-5) methyltransferase in human
RT   somatic tissues.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:9751-9756(1999).
RN   [6]
RP   PARTIAL SEQUENCE FROM N.A. (ISOFORM 2).
RX   MEDLINE=20219106; PubMed=10753866;
RA   Bonfils C., Beaulieu N., Chan E., Cotton-Montpetit J., MacLeod A.R.;
RT   "Characterization of the human DNA methyltransferase splice variant
RT   Dnmt1b.";
RL   J. Biol. Chem. 275:10754-10760(2000).
```

```
RN   [7]
RP   TISSUE SPECIFICITY.
RX   MEDLINE=99263031; PubMed=10325416;
RA   Robertson K.D., Uzvolgyi E., Liang G., Talmadge C., Sumegi J.,
RA   Gonzales F.A., Jones P.A.;
RT   "The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate
RT   mRNA expression in normal tissues and overexpression in tumors.";
RL   Nucleic Acids Res. 27:2291-2298(1999).
CC   -!- FUNCTION: Methylates CpG residues. Preferentially methylates
CC       hemimethylated DNA as its substrate. It is responsible for
CC       maintaining methylation patterns established in development.
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA = S-adenosyl-L-
CC       homocysteine + DNA containing 5-methylcytosine.
CC   -!- ALTERNATIVE PRODUCTS: 3 isoforms; 1 (shown here), 2/Dnmt1b and 3;
CC       are produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Ubiquitous. Isoform 2 is less expressed than
CC       isoform 1.
CC   -!- SIMILARITY: BELONGS TO THE C5-METHYLTRANSFERASE FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 BAH DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X63692; CAA45219.1; -.
DR   EMBL; AF180682; AAF23609.1; -.
DR   EMBL; AC010077; AAD54507.1; -.
DR   EMBL; AF169120; AAD51619.1; -.
DR   PIR; S19183; S19183.
DR   PIR; S22610; S22610.
DR   REBASE; 1161; M.HsaIA.
DR   MIM; 126375; -.
DR   InterPro; IPR001025; BAH.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   InterPro; IPR002857; Znf-CXXC.
DR   Pfam; PF01426; BAH; 2.
DR   Pfam; PF00145; DNA_methylase; 5.
DR   Pfam; PF02008; zf-CXXC; 2.
DR   PRINTS; PR00105; C5METTRFRASE.
DR   SMART; SM00439; BAH; 2.
DR   PROSITE; PS00094; C5_MTASE_1; 1.
DR   PROSITE; PS00095; C5_MTASE_2; 1.
KW   Transferase; Methyltransferase; DNA-binding; Repeat; Phosphorylation;
KW   Alternative splicing.
FT   DOMAIN      649    691       CYS/ARG/LYS-RICH.
FT   DOMAIN      755    880       BAH 1.
FT   DOMAIN      931   1100       BAH 2.
FT   ACT_SITE   1226   1226       BY SIMILARITY.
FT   MOD_RES     509    509       PHOSPHORYLATION (BY SIMILARITY).
FT   VARSPLIC      1    336       MISSING (IN ISOFORM 3).
FT   VARSPLIC    149    149       P -> RSRDPPASASQVTGIRA (IN ISOFORM 2).
SQ   SEQUENCE   1616 AA;  183164 MW;  1E833192D22AFA5B CRC64;


  Query Match             12.1%;  Score 73;  DB 1;  Length 1616;
  Best Local Similarity   26.7%;  Pred. No. 1.1e+02;
  Matches   32;  Conservative   18;  Mismatches   56;  Indels   14;  Gaps    5;

Qy       6 EEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           :|:   :|  |:|  :| |:     | || : :   | :  :    :   || :   |
Db     273 DEDEDGDEKDEKKHRSQPKDLAAKRRPEEKEPEKVNPQISDEKDEDEKEEKRRKTTPK-- 330

Qy      66 DSGDYNMAKAK--MKNKQLP----SAG---PDKNLVTGDHIPTPQDLPQRKSSLVTSKLA 116
             :  :  ||:||  | :|  |      |    | :|  | | |   | ||    |   ||:
Db     331 EPTEKKMARAKTVMNSKTHPPKCIQCGQYLDDPDLKYGQHPPDAVDEPQ---MLTNEKLS 387


RESULT  39
CABI_HUMAN
ID   CABI_HUMAN     STANDARD;      PRT;  2220 AA.
AC   Q9Y6J0; Q9Y460;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Calcineurin-binding protein Cabin 1 (Calcineurin inhibitor) (CAIN).
GN   CABIN1 OR KIAA0330.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=98318216; PubMed=9655484;
RA   Sun L., Youn H.-D., Loh C., Stolow M., He W., Liu J.O.;
RT   "Cabin 1, a negative regulator for calcineurin signaling in T
RT   lymphocytes.";
RL   Immunity 8:703-711(1998).
RN   [2]
RP   SEQUENCE OF 319-2220 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97349984; PubMed=9205841;
RA   Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 4:141-150(1997).
RN   [3]
RP   SEQUENCE OF 1545-2220 FROM N.A.
RX   MEDLINE=20057165; PubMed=10591208;
RA   Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA   Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.,
RA   Bagguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,
RA   Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,
RA   Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,
RA   Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,
RA   Coville G.J., Cox A.V., Davis J., Dawson E., Dhami P.D., Dockree C.,
RA   Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,
RA   Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E.,
RA   Grafham D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G.,
RA   Heathcott R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,
RA   Kimberley A., King A., Laird G.K., Langford C.F., Leversha M.A.,
RA   Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,
RA   Matthews L., Mccann O.T., Mcclay J., Mclaren S., Mcmurray A.A.,
RA   Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,
RA   Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,
RA   Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,
RA   Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,
RA   Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,
RA   Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,
RA   Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,
RA   Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,
RA   Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,
RA   Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,
RA   Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,
RA   Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E.,
RA   Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,
RA   Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.,
RA   Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Miller N.,
RA   Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,
RA   Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,
RA   Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,
RA   Rohlfing T., Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K.,
RA   Nelson J., Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R.,
RA   Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA   Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA   Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA   Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
RA   O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA   Khan A.S., Lane L., Tilahun Y., Wright H.;
RT   "The DNA sequence of human chromosome 22.";
```

```
RL   Nature 402:489-495(1999).
CC   -!- FUNCTION: IT MAY SERVE AS A NEGATIVE REGULATOR OF T CELL RECEPTOR
CC       (TCR) SIGNALING VIA INHIBITION OF CALCINEURIN. INTERACTS WITH AND
CC       INHIBITS CALCINEURIN-MEDIATED SIGNAL TRANSDUCTION. CABIN 1 IS
CC       SPECIFIC FOR THE ACTIVATED FORM OF CALCINEURIN, THEIR INTERACTION
CC       IS DEPENDENT ON BOTH PKC AND CALCIUM SIGNALS.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED IN DIFFERENT TISSUES.
CC   -!- PTM: ACTIVATED THROUGH PKC-MEDIATED HYPERPHOSPHORYLATION.
CC   -!- SIMILARITY: CONTAINS 6 TPR REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF072441; AAD40846.1; -.
DR   EMBL; AB002328; BAA20788.1; -.
DR   EMBL; Z92546; CAB62954.1; -.
DR   MIM; 604251; -.
DR   InterPro; IPR001440; TPR.
DR   Pfam; PF00515; TPR; 4.
DR   SMART; SM00028; TPR; 3.
KW   Phosphorylation; Repeat; TPR repeat.
FT   REPEAT       36     69       TPR 1.
FT   REPEAT       90    123       TPR 2.
FT   REPEAT      125    157       TPR 3.
FT   REPEAT      615    648       TPR 4.
FT   REPEAT     1055   1088       TPR 5.
FT   REPEAT     1106   1139       TPR 6.
FT   DOMAIN     2116   2153       CALCINEURIN BINDING (BY SIMILARITY).
SQ   SEQUENCE   2220 AA;  246350 MW;  BA4AD1741056C233 CRC64;


  Query Match             12.1%;  Score 73;  DB 1;  Length 2220;
  Best Local Similarity   29.7%;  Pred. No. 1.5e+02;
  Matches   27;  Conservative   10;  Mismatches   38;  Indels   16;  Gaps    3;

Qy      29 PERAEEAKLKAKYPSL------GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQL 82
           ||  :|   : | | |      | :|||   |:          |:||      : |:|:
Db    1696 PEDGQEGLPQPKKPPLADGSGPGPEPGGKVGLLNHRPVA---MDAGDSADQSGERKDKES 1752

Qy      83 PSAGPDKNLVT-------GDHIPTPQDLPQR 106
           | ||| : : |        |   ||  || |
Db    1753 PRAGPTEPMDTSEATVCHSDLERTPPLLPGR 1783


RESULT  40
PRGR_CHICK
ID   PRGR_CHICK     STANDARD;      PRT;   786 AA.
AC   P07812; Q90946;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Progesterone receptor (PR).
GN   PGR OR NR3C3.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88166640; PubMed=3443098;
RA   Gronemeyer H., Turcotte B., Quirin-Stricker C., Bocquel M.T.,
RA   Meyer M.E., Krozowski Z., Jeltsch J.M., Lerouge T., Garnier J.M.,
RA   Chambon P.;
RT   "The chicken progesterone receptor: sequence, expression and
RT   functional analysis.";
RL   EMBO J. 6:3985-3994(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91042592; PubMed=3153474;
RA   Conneely O.M., Dobson A.D.W., Tsai M.-J., Beattie W.G., Toft D.O.,
RA   Huckaby C.S., Zarucki T., Schrader W.T., O'Malley B.W.;
RT   "Sequence and expression of a functional chicken progesterone
RT   receptor.";
RL   Mol. Endocrinol. 1:517-525(1987).
RN   [3]
RP   SEQUENCE OF 128-164 FROM N.A.
RX   MEDLINE=86289413; PubMed=2426779;
RA   Conneely O.M., Sullivan W.P., Toft D.O., Birnbaumer M., Cook R.G.,
RA   Maxwell B.L., Zarucki-Schulz T., Greene G.L., Schrader W.T.,
RA   O'Malley B.W.;
RT   "Molecular cloning of the chicken progesterone receptor.";
RL   Science 233:767-770(1986).
RN   [4]
RP   SEQUENCE OF 417-490 FROM N.A.
RX   MEDLINE=86287271; PubMed=2426697;
RA   Jeltsch J.M., Krozowski Z., Quirin-Stricker C., Gronemeyer H.,
RA   Simpson R.J., Garnier J.M., Krust A., Jacob F., Chambon P.;
RT   "Cloning of the chicken progesterone receptor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:5424-5428(1986).
RN   [5]
RP   DIFFERENCE BETWEEN FORM 1 AND FORM 2.
RX   MEDLINE=89340509; PubMed=2760059;
RA   Conneely O.M., Kettelberger D.M., Tsai M.-J., Schrader W.T.,
RA   O'Malley B.W.;
RT   "The chicken progesterone receptor A and B isoforms are products of
RT   an alternate translation initiation event.";
RL   J. Biol. Chem. 264:14062-14064(1989).
RN   [6]
RP   SEQUENCE FROM N.A. (ALL FORMS).
RX   MEDLINE=90154085; PubMed=2303488;
RA   Jeltsch J.-M., Turcotte B., Garnier J.-M., Lerouge T., Krozowski Z.,
RA   Gronemeyer H., Chambon P.;
RT   "Characterization of multiple mRNAs originating from the chicken
RT   progesterone receptor gene. Evidence for a specific transcript
RT   encoding form A.";
RL   J. Biol. Chem. 265:3967-3974(1990).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS; A (SHOWN HERE), A', B AND B';
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: OVIDUCT.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y00092; CAA68282.1; -.
DR   EMBL; M13972; AAA49034.1; -.
DR   EMBL; M37518; AAA49013.1; -.
DR   EMBL; M37518; AAA49014.1; -.
DR   EMBL; M14278; AAA49035.1; -.
DR   EMBL; M14279; AAA49038.1; -.
DR   EMBL; M14280; AAA49039.1; -.
DR   EMBL; M32732; AAA49011.1; -.
DR   EMBL; M31104; AAA49011.1; JOINED.
DR   EMBL; M32726; AAA49011.1; JOINED.
DR   EMBL; M32727; AAA49011.1; JOINED.
DR   EMBL; M32728; AAA49011.1; JOINED.
DR   EMBL; M32729; AAA49011.1; JOINED.
```

```
DR   EMBL; M32730; AAA49011.1; JOINED.
DR   EMBL; M32732; AAA49012.1; -.
DR   EMBL; M31104; AAA49012.1; JOINED.
DR   EMBL; M32726; AAA49012.1; JOINED.
DR   EMBL; M32727; AAA49012.1; JOINED.
DR   EMBL; M32728; AAA49012.1; JOINED.
DR   EMBL; M32729; AAA49012.1; JOINED.
DR   EMBL; M32730; AAA49012.1; JOINED.
DR   EMBL; M31104; AAA49009.1; -.
DR   EMBL; M31104; AAA49010.1; -.
DR   PIR; A24312; A24312.
DR   PIR; A24661; A24661.
DR   PIR; S06284; S06284.
DR   HSSP; P06401; 1A28.
DR   TRANSFAC; T00698; -.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR000128; Progest_receptor.
DR   InterPro; IPR001628; zf-C4.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF02161; Prog_receptor; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding; Alternative splicing.
FT   DOMAIN        1    420       MODULATING, PRO-RICH.
FT   DNA_BIND    421    486       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     421    441       C4-TYPE.
FT   ZN_FING     457    481       C4-TYPE.
FT   DOMAIN      487    786       STEROID-BINDING.
FT   DOMAIN       48     80       ASP/GLU-RICH (ACIDIC).
FT   VARSPLIC      1    127       MISSING (IN ISOFORM B AND ISOFORM B').
FT   VARSPLIC    452    458       QHNYLCA -> TISYHCS (IN ISOFORM A' AND
FT                                ISOFORM B').
FT   VARSPLIC    459    786       MISSING (IN ISOFORM A' AND ISOFORM B').
FT   CONFLICT     58     58       E -> DD (IN REF. 2).
FT   CONFLICT    480    480       K -> N (IN REF. 2).
FT   CONFLICT    489    489       G -> A (IN REF. 2).
FT   CONFLICT    577    577       R -> T (IN REF. 2).
FT   CONFLICT    642    642       M -> I (IN REF. 2).
SQ   SEQUENCE   786 AA;  85743 MW;  659559950BC45ED9 CRC64;


  Query Match             12.0%;  Score 72.5;  DB 1;  Length 786;
  Best Local Similarity   27.4%;  Pred. No. 56;
  Matches   29;  Conservative   14;  Mismatches   46;  Indels   17;  Gaps    4;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQ 62
           :::|||| ||  || |  :|:|    |  || :    |     :|||       |
Db      50 EEEEEENEEEEEEEEPQQREEEE----EEEEEDRDCPSY-----RPGGGSLSKDCLDSVL 100

Qy      63 KYFDSGDYNMAKAKMKNKQLP-------SAGPDKNLV-TGDHIPTP 100
             | :   : |   :   ::|       | ||::  |  |   | |
Db     101 DTFLAPAAHAAPWSLFGPEVPEVPVAPMSRGPEQKAVDAGPGAPGP 146


Search completed: September  7, 2002, 10:51:16
Job time: 392 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2002, 15:34:27 ; Search time 4285.42 Seconds
(without alignments)
5854.949 Million cell updates/sec

Title: US-09-824-178-2
Perfect score: 1199
Sequence: 1 GTGACAGGAGCCGAAGCAGC.....CCAAAAAAAAAAAAAAAAAA 1199

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1149	95.8	1199	9	AF157509	AF157509 Homo sapi
2	1135.8	94.7	1211	9	BC000436	BC000436 Homo sapi
3	1135	94.7	1202	9	BC004461	BC004461 Homo sapi
4	1131.4	94.4	1208	9	AK001981	AK001981 Homo sapi
c 5	678.4	56.6	176850	9	AL356356	AL356356 Human DNA
c 6	678.4	56.6	207815	2	AC053497	AC053497 Homo sapi
7	666	55.5	749	9	HSALPEND	X99906 Homo sapien
8	613	51.1	1788	9	HSAJ10966	AJ010966 Homo sapi
9	501.8	41.9	669	6	AR144151	AR144151 Sequence
10	501.8	41.9	669	6	AR176338	AR176338 Sequence
11	456	38.0	1093	9	AF157510	AF157510 Homo sapi
c 12	434.4	36.2	148578	9	CNS01RIF	AL163171 Human chr
c 13	434.4	36.2	159846	2	AC069588	AC069588 Homo sapi
14	383.6	32.0	176616	2	AC007449	AC007449 Homo sapi
c 15	381.6	31.8	89263	9	AC007124	AC007124 Human gen
c 16	380	31.7	89872	9	HSJ822J19	AL050324 Human DNA
17	350	29.2	366	4	SSA005987	AJ005987 Sus scrof
18	348.4	29.1	366	4	BTA005986	AJ005986 Bos tauru
19	348.4	29.1	366	10	MMA005985	AJ005985 Mus muscu
20	345.2	28.8	366	10	RNA005984	AJ005984 Rattus no
21	287.4	24.0	194304	9	AC016708	AC016708 Homo sapi
22	286.4	23.9	98113	2	AC023294	AC023294 Homo sapi
23	229.4	19.1	69732	2	AC091596	AC091596 Homo sapi
24	213.4	17.8	231	4	BTAENSULF	X95958 B.taurus mR
c 25	185	15.4	142023	2	AC092479	AC092479 Mus muscu
26	175.4	14.6	339	10	MMA005983	AJ005983 Mus muscu
27	172	14.3	105126	2	AC103421	AC103421 Rattus no
28	170.2	14.2	339	10	RNA005982	AJ005982 Rattus no
29	168.6	14.1	339	9	HOS223091	AJ223091 Homo sapi
30	168.6	14.1	1188	4	BOVARPP19	M33618 Bovine cAMP
31	168.6	14.1	1228	9	BC003418	BC003418 Homo sapi
32	168.6	14.1	5171	9	AF084555	AF084555 Homo sapi
33	167.2	13.9	71941	2	AC091874	AC091874 Homo sapi
34	167	13.9	339	4	SSA005981	AJ005981 Sus scrof
35	165.4	13.8	2288	9	AK057673	AK057673 Homo sapi
36	158.6	13.2	1961	4	BOVARPP16	M33617 Bovine cAMP
37	157.2	13.1	173169	9	AC068538	AC068538 Homo sapi
c 38	148.8	12.4	170548	10	AC074041	AC074041 Mus muscu
c 39	99.2	8.3	70244	2	AC090836	AC090836 Homo sapi
c 40	99.2	8.3	149081	2	AC025917	AC025917 Homo sapi
c 41	99.2	8.3	184128	2	AC073438	AC073438 Homo sapi
c 42	99.2	8.3	219389	2	AC009754	AC009754 Homo sapi
43	80.2	6.7	221899	2	AC079548	AC079548 Mus muscu
44	74	6.2	246262	2	AL669825	AL669825 Mus muscu
45	71.6	6.0	194799	2	AC096386	AC096386 Rattus no

ALIGNMENTS

RESULT 1
AF157509
LOCUS AF157509 1199 bp mRNA linear PRI 01-JUL-2000
DEFINITION Homo sapiens alpha endosulfine mRNA, complete cds.
ACCESSION AF157509
VERSION AF157509.1 GI:8886008
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1199)
AUTHORS Scott,V.E.S., Roch,J.-M., Davis-Taber,R.A., Molinari,E.J., Whiteaker,K.L., Gopalakrishnan,M., Idler,K. and Sullivan,J.P.
TITLE Cloning and molecular characterization of two isoforms of human endosulfine
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1199)
AUTHORS Scott,V.E.S., Roch,J.-M., Whiteaker,K.L. and Sullivan,J.P.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-1999) D47C, Abbott Laboratories, AP10 L08, 100 Abbott Park, Abbott Park, IL 60064, USA
FEATURES Location/Qualifiers

```
     source          1. .1199
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             107. .472
                     /codon_start=1
                     /product="alpha endosulfine"
                     /protein_id="AAF80340.1"
                     /db_xref="GI:8886009"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAGGQVE"
BASE COUNT      335 a    288 c    333 g    243 t
ORIGIN
```

```
  Query Match             95.8%;  Score 1149;  DB 9;  Length 1199;
  Best Local Similarity   98.3%;  Pred. No. 0;
  Matches 1189;  Conservative    0;  Mismatches    0;  Indels   20;  Gaps    2;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
Db       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
Db      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
Db     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
Db     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
Db     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
Db     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
Db     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480
Db     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480

Qy     481 GGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTC 540
Db     481 GGGGCTCTGCCAGATCCTGAGA----------CGCTTCCCCTCCCTGCCCCACCCGGGTC 530

Qy     541 CTGTGCTGGCTCCTGCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTG 590
Db     531 CTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTG 590

Qy     591 TGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGC 650
Db     591 TGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGC 650

Qy     651 TGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGA 710
Db     651 TGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGA 710

Qy     711 AATATCAATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGGA 770
Db     711 AATATCAATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGGA 770

Qy     771 GGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACG 830
Db     771 GGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACG 830

Qy     831 TGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTC 890
Db     831 TGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTC 890

Qy     891 TAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATA 950
Db     891 TAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATA 950

Qy     951 AAAGAAACTCCATCCGCACAGACAGATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAGTT 1010
Db     951 AAAGAAACTCCATCCGCACAGACAGATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAGTT 1010

Qy    1011 GAGAACTCAACATTAATCATTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGGGAA 1070
Db    1011 GAGAACTCAACATTAATCATTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGGGAA 1070

Qy    1071 GAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAA 1130
Db    1071 GAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAA 1130

Qy    1131 GGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAA 1190
Db    1131 GGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAA 1190

Qy    1191 AAAAAAAAA 1199
Db    1191 AAAAAAAAA 1199
```

```
RESULT   2
BC000436
LOCUS       BC000436               1211 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, endosulfine alpha, clone MGC:8394 IMAGE:2820528,
            mRNA, complete cds.
ACCESSION   BC000436
VERSION     BC000436.1  GI:12653334
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1211)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site:       http://www.nisc.nih.gov/
            Contact:        nisc_mgc@nhgri.nih.gov
            Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R.,
            Lim,M., Maduro,Q.L., Masiello,C., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Snyder,B., Stantripop,S., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 1 Row: d Column: 1
            This clone was selected for full length sequencing because it
```

```
          passed the following selection criteria: matched mRNA gi: 8886008.
FEATURES             Location/Qualifiers
     source          1. .1211
                     /organism="Homo sapiens"
                     /db_xref="LocusID:2029"
                     /db_xref="taxon:9606"
                     /clone="MGC:8394 IMAGE:2820528"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             110. .475
                     /codon_start=1
                     /product="endosulfine alpha"
                     /protein_id="AAH00436.1"
                     /db_xref="GI:12653335"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAGGQVE"
BASE COUNT      340 a    291 c    337 g    243 t
ORIGIN


  Query Match             94.7%;  Score 1135.8;  DB 9;  Length 1211;
  Best Local Similarity   97.8%;  Pred. No. 0;
  Matches 1180;  Conservative    0;  Mismatches    7;  Indels   20;  Gaps    2;

Qy      3 GACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCG 62
          ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6 GAGGGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCG 65

Qy     63 GTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAG 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66 GTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAG 125

Qy    123 AAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAG 182
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    126 AAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAG 185

Qy    183 GTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGAC 242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    186 GTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGAC 245

Qy    243 AAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTG 302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    246 AAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTG 305

Qy    303 ACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAG 362
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    306 ACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAG 365

Qy    363 GACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGA 422
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    366 GACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGA 425

Qy    423 GAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGG 482
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    426 GAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGG 485

Qy    483 GGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCT 542
          ||||||||||||||||||||          ||||||||||||||||||||||||||||||
Db    486 GGCTCTGCCAGATCCTGAGA----------CGCTTCCCCTCCCTGCCCCACCCGGGTCCT 535

Qy    543 GTGCTGGCTCCTGCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTGTG 592
          ||||||||||||||||||          ||||||||||||||||||||||||||||||||
Db    536 GTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTG 595

Qy    593 GGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTG 652
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    596 GGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTG 655

Qy    653 AGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAA 712
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    656 AGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAA 715

Qy    713 TATCAATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGGAGG 772
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    716 TATCAATGAAATTCTGGGGGTGGGGGTGGGGTTGAAAGGGTGGGGTGGGAGATATGGAGG 775

Qy    773 AGTATGAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACGTG 832
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    776 AGTATGAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACGTG 835

Qy    833 GCTGATGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTA 892
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    836 GCTGATGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTA 895

Qy    893 GTTGAGCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATAAA 952
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    896 GTTGAGCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATAAA 955

Qy    953 AGAAACTCCATCCGCACAGACAGATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAGTTGA 1012
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    956 AGAAACTCCATCCGCACAGACAGATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAGTTGA 1015

Qy   1013 GAACTCAACATTAATCATTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGGGAAGA 1072
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1016 GAACTCAACATTAATCATTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGGGAAGA 1075

Qy   1073 AGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGG 1132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1076 AGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGG 1135

Qy   1133 ACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAA 1192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||    | ||||
Db   1136 ACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAACCCAGAAAA 1195

Qy   1193 AAAAAAA 1199
          |||||||
Db   1196 AAAAAAA 1202


RESULT   3
BC004461
LOCUS       BC004461                1202 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, endosulfine alpha, clone MGC:4319 IMAGE:2820528,
            mRNA, complete cds.
ACCESSION   BC004461
VERSION     BC004461.1  GI:13325294
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1202)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
```

Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 9 Row: a Column: 17
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8886008.

FEATURES Location/Qualifiers
source 1. .1202
/organism="Homo sapiens"
/db_xref="LocusID:2029"
/db_xref="taxon:9606"
/clone="MGC:4319 IMAGE:2820528"
/tissue_type="Lung, small cell carcinoma"
/clone_lib="NIH_MGC_7"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
CDS 101. .466
/codon_start=1
/product="endosulfine alpha"
/protein_id="AAH04461.1"
/db_xref="GI:13325295"
/translation="MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD LPQRKSSLVTSKLAGGQVE"
BASE COUNT 338 a 289 c 332 g 243 t
ORIGIN

Query Match 94.7%; Score 1135; DB 9; Length 1202;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 1178; Conservative 0; Mismatches 5; Indels 20; Gaps 2;

```
Qy       7 GGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTG 66
Db       1 GGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTG 60

Qy      67 CCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGA 126
Db      61 CCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGA 120

Qy     127 AGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTAT 186
Db     121 AGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTAT 180

Qy     187 TCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAA 246
Db     181 TCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAA 240

Qy     247 GCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTC 306
Db     241 GCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTC 300

Qy     307 AGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACC 366
Db     301 AGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACC 360

Qy     367 AGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAA 426
Db     361 AGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAA 420

Qy     427 GTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCT 486
Db     421 GTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCT 480

Qy     487 CTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGC 546
Db     481 CTGCCAGATCCTGAGA----------CGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGC 530

Qy     547 TGGCTCCTGCCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCT 596
Db     531 TGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCT 590

Qy     597 GGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCA 656
Db     591 GGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCA 650

Qy     657 CCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATC 716
Db     651 CCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATC 710

Qy     717 AATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGGAGGAGTA 776
Db     711 AATGAAATTCTGGGGGTGGGGGTGGGGTTGAAAGGGTGGGGTGGGAGATATGGAGGAGTA 770

Qy     777 TGAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACGTGGCTG 836
Db     771 TGAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACGTGGCTG 830

Qy     837 ATGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTAGTTG 896
Db     831 ATGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTAGTTG 890

Qy     897 AGCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATAAAAGAA 956
Db     891 AGCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATAAAAGAA 950

Qy     957 ACTCCATCCGCACAGACAGATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAGTTGAGAAC 1016
Db     951 ACTCCATCCGCACAGACAGATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAGTTGAGAAC 1010

Qy    1017 TCAACATTAATCATTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGGGAAGAAGGG 1076
Db    1011 TCAACATTAATCATTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGGGAAGAAGGG 1070

Qy    1077 CAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTT 1136
Db    1071 CAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTT 1130

Qy    1137 TTTCAGTATTCGAAATAAACAAAACAAAAAACCCATGAAAAAACCCAAAAAAAAAAAAAAA 1196
Db    1131 TTTCAGTATTCGAAATAAACAAAACAAAAAACCCATGAAAAAACCCAACCCAGAAAAAAAA 1190

Qy    1197 AAA 1199
Db    1191 AAA 1193
```

RESULT 4
AK001981
LOCUS AK001981 1208 bp mRNA linear PRI 22-FEB-2000
DEFINITION Homo sapiens cDNA FLJ11119 fis, clone PLACE1006040, highly similar to Homo sapiens mRNA for alpha endosulfine.
ACCESSION AK001981
VERSION AK001981.1 GI:7023587
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens placenta cDNA to mRNA, clone_lib:PLACE1 clone:PLACE1006040.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M., Hiraoka,S., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Kanehori,K.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 1208)
AUTHORS Isogai,T. and Otsuki,T.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,

```
            Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1. .1208
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PLACE1006040"
                     /clone_lib="PLACE1"
                     /tissue_type="placenta"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      326 a    295 c    337 g    250 t
ORIGIN


  Query Match             94.4%;  Score 1131.4;  DB 9;  Length 1208;
  Best Local Similarity   98.2%;  Pred. No. 0;
  Matches 1172;  Conservative    0;  Mismatches    1;  Indels   20;  Gaps    2;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 80

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 140

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 200

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 260

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 320

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 380

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 440

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     441 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 500

Qy     481 GGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTC 540
           ||||||||||||||||||||          ||||||||||||||||||||||||||||||
Db     501 GGGGCTCTGCCAGATCCTGAGA----------CGCTTCCCCTCCCTGCCCCACCCGGGTC 550

Qy     541 CTGTGCTGGCTCCTGCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTG 590
           ||||||||||||||||||||          ||||||||||||||||||||||||||||||
Db     551 CTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTG 610

Qy     591 TGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGC 650
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     611 TGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGC 670

Qy     651 TGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGA 710
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     671 TGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGA 730

Qy     711 AATATCAATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGGA 770
           ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db     731 AATATCAATGAAATTCTGGGGGTGGGGGTGGGGTTGAAAGGGTGGGGTGGGAGATATGGA 790

Qy     771 GGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACG 830
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     791 GGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACG 850

Qy     831 TGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTC 890
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     851 TGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTC 910

Qy     891 TAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATA 950
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     911 TAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATA 970

Qy     951 AAAGAAACTCCATCCGCACAGACAGATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAGTT 1010
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     971 AAAGAAACTCCATCCGCACAGACAGATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAGTT 1030

Qy    1011 GAGAACTCAACATTAATCATTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGGGAA 1070
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1031 GAGAACTCAACATTAATCATTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGGGAA 1090

Qy    1071 GAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAA 1130
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1091 GAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAA 1150

Qy    1131 GGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAA 1183
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1151 GGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAA 1203


RESULT   5
AL356356/c
LOCUS       AL356356              176550 bp    DNA     linear   PRI 30-JAN-2002
DEFINITION  Human DNA sequence from clone RP11-54A4 on chromosome 1, complete
            sequence.
ACCESSION   AL356356
VERSION     AL356356.17  GI:18476570
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176550)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 1, 2002 this sequence version replaced gi:18250748.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
```

```
            RP11-54A4 is from the library RPCI-11.1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-54A4 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-54A4 is at 1 in this sequence. The
            true left end of clone RP11-363I22 is at 174551 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .176550
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-54A4"
                     /clone_lib="RPCI-11.1"
     misc_feature    50307. .50323
                     /note="Single clone region. Reads generated from a
                     transposon library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    123022. .123038
                     /note="Sequence from uni-directional primer reads and dGTP
                     big dye terminator reads only."
BASE COUNT    45999 a  42218 c  42262 g  46071 t
ORIGIN


  Query Match             56.6%;  Score 678.4;  DB 9;  Length 176550;
  Best Local Similarity   96.5%;  Pred. No. 4.2e-181;
  Matches  722;  Conservative    0;  Mismatches    6;  Indels   20;  Gaps    2;

Qy      456 GTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGC 515
            ||||||||||||||||||||||||||||||||||||||||||||||||||          |||
Db   150122 GTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGA----------CGC 150073

Qy      516 TTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTT----------CAGC 565
            ||||||||||||||||||||||||||||||||||||||||||||||          ||||
Db   150072 TTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGC 150013

Qy      566 CAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTG 625
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   150012 CAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTG 149953

Qy      626 TCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACC 685
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   149952 TCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACC 149893

Qy      686 TTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGTGGGGCT 745
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   149892 TTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGTGGGGTT 149833

Qy      746 GAAAGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACA 805
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   149832 GAAAGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACA 149773

Qy      806 TTCCTGACTATCCTTCTTAACCACGTGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAGT 865
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   149772 TTCCTGACTATCCTTCTTAACCACGTGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAGT 149713

Qy      866 GGAGTAGCCTAATGAAAAGGGGTTCTAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAGT 925
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   149712 GGAGTAGCCTAATGAAAAGGGGTTCTAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAGT 149653

Qy      926 GTGGTTGGAGACCTGGTGTCAGATAAAAGAAACTCCATCCGCACAGACAGATGCAAACAG 985
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   149652 GTGGTTGGAGACCTGGTGTCAGATAAAAGAAACTCCATCCGCACAGACAGATGCAAACAG 149593

Qy      986 CTCCTCTAGTTCTGCAGAGCTAGTTGAGAACTCAACATTAATCATTTTAAAAAGTACTGT 1045
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   149592 CTCCTCTAGTTCTGCAGAGCTAGTTGAGAACTCAACATTAATCATTTTAAAAAGTACTGT 149533

Qy     1046 CCTTGAAATAGATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAG 1105
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   149532 CCTTGAAATAGATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAG 149473

Qy     1106 CGTGGGGAACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAA 1165
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   149472 CGTGGGGAACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAA 149413

Qy     1166 ACCCATGAAAAAACCCAAAAAAAAAAAA 1193
            |||||||||||||||||    | ||| |
Db   149412 ACCCATGAAAAAACCCAACCCAGAAATA 149385


RESULT   6
AC053497/c
LOCUS       AC053497              207815 bp    DNA     linear   HTG 13-AUG-2000
DEFINITION  Homo sapiens chromosome 1 clone RP11-243G22, WORKING DRAFT
            SEQUENCE, 37 unordered pieces.
ACCESSION   AC053497
VERSION     AC053497.4  GI:9797901
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 207815)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 207815)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8439979.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0243G22
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 192323 bases at least Q40
            Consensus quality: 196320 bases at least Q30
            Consensus quality: 198661 bases at least Q20
            Insert size: 204000; agarose-fp
            Insert size: 204215; sum-of-contigs
            Quality coverage: 6.36 in Q20 bases; agarose-fp
            Quality coverage: 6.31 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 37 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1381: contig of 1381 bp in length
            *     1382     1481: gap of unknown length
            *     1482     2602: contig of 1121 bp in length
            *     2603     2702: gap of unknown length
            *     2703     4013: contig of 1311 bp in length
            *     4014     4113: gap of unknown length
            *     4114     5331: contig of 1218 bp in length
            *     5332     5431: gap of unknown length
            *     5432     6622: contig of 1191 bp in length
            *     6623     6722: gap of unknown length
```

```
            *     6723     8166: contig of 1444 bp in length
            *     8167     8266: gap of unknown length
            *     8267     9674: contig of 1408 bp in length
            *     9675     9774: gap of unknown length
            *     9775    10943: contig of 1169 bp in length
            *    10944    11043: gap of unknown length
            *    11044    12198: contig of 1155 bp in length
            *    12199    12298: gap of unknown length
            *    12299    14193: contig of 1895 bp in length
            *    14194    14293: gap of unknown length
            *    14294    15717: contig of 1424 bp in length
            *    15718    15817: gap of unknown length
            *    15818    16935: contig of 1118 bp in length
            *    16936    17035: gap of unknown length
            *    17036    18375: contig of 1340 bp in length
            *    18376    18475: gap of unknown length
            *    18476    19962: contig of 1487 bp in length
            *    19963    20062: gap of unknown length
            *    20063    21169: contig of 1107 bp in length
            *    21170    21269: gap of unknown length
            *    21270    22452: contig of 1183 bp in length
            *    22453    22552: gap of unknown length
            *    22553    24582: contig of 2030 bp in length
            *    24583    24682: gap of unknown length
            *    24683    26383: contig of 1701 bp in length
            *    26384    26483: gap of unknown length
            *    26484    27710: contig of 1227 bp in length
            *    27711    27810: gap of unknown length
            *    27811    29195: contig of 1385 bp in length
            *    29196    29295: gap of unknown length
            *    29296    30800: contig of 1505 bp in length
            *    30801    30900: gap of unknown length
            *    30901    32555: contig of 1655 bp in length
            *    32556    32655: gap of unknown length
            *    32656    34653: contig of 1998 bp in length
            *    34654    34753: gap of unknown length
            *    34754    37122: contig of 2369 bp in length
            *    37123    37222: gap of unknown length
            *    37223    39506: contig of 2284 bp in length
            *    39507    39606: gap of unknown length
            *    39607    41220: contig of 1614 bp in length
            *    41221    41320: gap of unknown length
            *    41321    42978: contig of 1658 bp in length
            *    42979    43078: gap of unknown length
            *    43079    45773: contig of 2695 bp in length
            *    45774    45873: gap of unknown length
            *    45874    55056: contig of 9183 bp in length
            *    55057    55156: gap of unknown length
            *    55157    67059: contig of 11903 bp in length
            *    67060    67159: gap of unknown length
            *    67160    80260: contig of 13101 bp in length
            *    80261    80360: gap of unknown length
            *    80361    96528: contig of 16168 bp in length
            *    96529    96628: gap of unknown length
            *    96629   114529: contig of 17901 bp in length
            *   114530   114629: gap of unknown length
            *   114630   133673: contig of 19044 bp in length
            *   133674   133773: gap of unknown length
            *   133774   153791: contig of 20018 bp in length
            *   153792   153891: gap of unknown length
            *   153892   176079: contig of 22188 bp in length
            *   176080   176179: gap of unknown length
            *   176180   207815: contig of 31636 bp in length.
FEATURES             Location/Qualifiers
     source          1. .207815
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-243G22"
     misc_feature    1. .1381
                     /note="assembly_name:Contig13"
     misc_feature    1482. .2602
                     /note="assembly_name:Contig15"
     misc_feature    2703. .4013
                     /note="assembly_name:Contig21"
     misc_feature    4114. .5331
                     /note="assembly_name:Contig25"
     misc_feature    5432. .6622
                     /note="assembly_name:Contig33"
     misc_feature    6723. .8166
                     /note="assembly_name:Contig37"
     misc_feature    8267. .9674
                     /note="assembly_name:Contig38"
     misc_feature    9775. .10943
                     /note="assembly_name:Contig40"
     misc_feature    11044. .12198
                     /note="assembly_name:Contig42"
     misc_feature    12299. .14193
                     /note="assembly_name:Contig45"
     misc_feature    14294. .15717
                     /note="assembly_name:Contig47"
     misc_feature    15818. .16935
                     /note="assembly_name:Contig50"
     misc_feature    17036. .18375
                     /note="assembly_name:Contig51"
     misc_feature    18476. .19962
                     /note="assembly_name:Contig52"
     misc_feature    20063. .21169
                     /note="assembly_name:Contig53"
     misc_feature    21270. .22452
                     /note="assembly_name:Contig54"
     misc_feature    22553. .24582
                     /note="assembly_name:Contig56"
     misc_feature    24683. .26383
                     /note="assembly_name:Contig57"
     misc_feature    26484. .27710
                     /note="assembly_name:Contig58"
     misc_feature    27811. .29195
                     /note="assembly_name:Contig59"
     misc_feature    29296. .30800
                     /note="assembly_name:Contig60"
     misc_feature    30901. .32555
                     /note="assembly_name:Contig63"
     misc_feature    32656. .34653
                     /note="assembly_name:Contig64"
     misc_feature    34754. .37122
                     /note="assembly_name:Contig65"
     misc_feature    37223. .39506
                     /note="assembly_name:Contig66"
     misc_feature    39607. .41220
                     /note="assembly_name:Contig67"
     misc_feature    41321. .42978
                     /note="assembly_name:Contig68"
     misc_feature    43079. .45773
                     /note="assembly_name:Contig69
                     clone_end:T7
                     vector_side:left"
     misc_feature    45874. .55056
                     /note="assembly_name:Contig70"
     misc_feature    55157. .67059
                     /note="assembly_name:Contig71"
     misc_feature    67160. .80260
                     /note="assembly_name:Contig72"
     misc_feature    80361. .96528
                     /note="assembly_name:Contig73"
     misc_feature    96629. .114529
                     /note="assembly_name:Contig74"
     misc_feature    114630. .133673
                     /note="assembly_name:Contig75"
     misc_feature    133774. .153791

  Query Match             56.6%;  Score 678.4;  DB 2;  Length 207815;
  Best Local Similarity   96.5%;  Pred. No. 4.3e-181;
  Matches  722;  Conservative    0;  Mismatches    6;  Indels   20;  Gaps    2;

Qy    456 GTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGC 515
```

```
Db  89143 GTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGA----------CGC 89094

Qy    516 TTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTT----------CAGC 565

Db  89093 TTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGC 89034

Qy    566 CAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTG 625

Db  89033 CAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTG 88974

Qy    626 TCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACC 685

Db  88973 TCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACC 88914

Qy    686 TTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGTGGGGCT 745

Db  88913 TTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGTGGGGTT 88854

Qy    746 GAAAGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACA 805

Db  88853 GAAAGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACA 88794

Qy    806 TTCCTGACTATCCTTCTTAACCACGTGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAGT 865

Db  88793 TTCCTGACTATCCTTCTTAACCACGTGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAGT 88734

Qy    866 GGAGTAGCCTAATGAAAAGGGGTTCTAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAGT 925

Db  88733 GGAGTAGCCTAATGAAAAGGGGTTCTAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAGT 88674

Qy    926 GTGGTTGGAGACCTGGTGTCAGATAAAAGAAACTCCATCCGCACAGACAGATGCAAACAG 985

Db  88673 GTGGTTGGAGACCTGGTGTCAGATAAAAGAAACTCCATCCGCACAGACAGATGCAAACAG 88614

Qy    986 CTCCTCTAGTTCTGCAGAGCTAGTTGAGAACTCAACATTAATCATTTTAAAAAGTACTGT 1045

Db  88613 CTCCTCTAGTTCTGCAGAGCTAGTTGAGAACTCAACATTAATCATTTTAAAAAGTACTGT 88554

Qy   1046 CCTTGAAATAGATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAG 1105

Db  88553 CCTTGAAATAGATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAG 88494

Qy   1106 CGTGGGGAACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAA 1165

Db  88493 CGTGGGGAACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAA 88434

Qy   1166 ACCCATGAAAAAACCCAAAAAAAAAAAA 1193

Db  88433 ACCCATGAAAAAACCCAACCCAGAAATA 88406


RESULT   7
HSALPEND
LOCUS       HSALPEND                749 bp    mRNA    linear   PRI 22-SEP-1998
DEFINITION  Homo sapiens mRNA for alpha endosulfine.
ACCESSION   X99906
VERSION     X99906.1  GI:2764973
KEYWORDS    alpha endosulfine; sulfonylurea receptor.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 749)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-1996) D. Bataille, Inserm U 376, Endocrinologie
            des Peptides, 371, Rue du Doyen Gaston Giraud, Montpellier Cedex
            05, 34294, FRANCE
REFERENCE   2  (bases 1 to 749)
  AUTHORS   Heron,L., Virsolvy,A., Peyrollier,K., Gribble,F.M., Le Cam,A.,
            Ashcroft,F.M. and Bataille,D.
  TITLE     Human alpha-endosulfine, a possible regulator of
            sulfonylurea-sensitive KATP channel: molecular cloning, expression
            and biological properties
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (14), 8387-8391 (1998)
  MEDLINE   98318667
FEATURES             Location/Qualifiers
     source          1. .749
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="brain"
                     /dev_stage="adult"
     CDS             126. .491
                     /function="endogenous lignad for sulfonylurea receptor"
                     /codon_start=1
                     /product="alpha endosulfine"
                     /protein_id="CAA68180.1"
                     /db_xref="GI:2764974"
                     /db_xref="SPTREMBL:O43768"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAGGQVE"
BASE COUNT      184 a    224 c    200 g    141 t
ORIGIN


  Query Match           55.5%;  Score 666;  DB 9;  Length 749;
  Best Local Similarity  97.2%;  Pred. No. 4.7e-178;
  Matches  717;  Conservative    0;  Mismatches    0;  Indels   21;  Gaps    3;

Qy      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60

Db     20 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 79

Qy     61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120

Db     80 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 139

Qy    121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180

Db    140 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 199

Qy    181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240

Db    200 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 259

Qy    241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300

Db    260 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 319

Qy    301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360

Db    320 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 379

Qy    361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420

Db    380 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 439

Qy    421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480

Db    440 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 499

Qy    481 GGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTC 540

Db    500 GGGGCTCTGCCAGATCCTGAGA----------CGCTTCCCCTCCCTGCCCCACCCGGGTC 549

Qy    541 CTGTGCTGGCTCCTGCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTG 590

Db    550 CTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTG 609

Qy    591 TGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGC 650

Db    610 TGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGA-CCCCTTGGGC 668

Qy    651 TGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGA 710
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    669 TGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGA 728

Qy    711 AATATCAATGAAATTCTG 728
          ||||||||||||||||||
Db    729 AATATCAATGAAATTCTG 746


RESULT   8
HSAJ10966
LOCUS       HSAJ10966               1788 bp    DNA     linear   PRI 07-JAN-2000
DEFINITION  Homo sapiens ENSA gene.
ACCESSION   AJ010966
VERSION     AJ010966.1  GI:6688263
KEYWORDS    alpha endosulfine; ENSA gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1788)
  AUTHORS   Heron,L., Virsolvy,A., Peyrollier,K., Gribble,F.M., Le Cam,A.,
            Ashcroft,F.M. and Bataille,D.
  TITLE     Human alpha-endosulfine, a possible regulator of
            sulfonylurea-sensitive KATP channel: molecular cloning, expression
            and biological properties
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (14), 8387-8391 (1998)
  MEDLINE   98318667
REFERENCE   2  (bases 1 to 1788)
  AUTHORS   Bataille,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1788)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-1998) Bataille D., INSERM, U 376, CHU
            Arnaud-de-Villeneuve, Montpellier, 34295, FRANCE
  REMARK    revised by author
FEATURES             Location/Qualifiers
     source          1. .1788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            757. .1122
                     /gene="ENSA"
     CDS             757. .1122
                     /gene="ENSA"
                     /function="putative regulator of K-ATP channels"
                     /codon_start=1
                     /evidence=experimental
                     /product="alpha endosulfine"
                     /protein_id="CAB65125.1"
                     /db_xref="GI:6688264"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAGGQVE"
BASE COUNT      532 a    388 c    411 g    457 t
ORIGIN


  Query Match           51.1%;  Score 613;  DB 9;  Length 1788;
  Best Local Similarity   96.2%;  Pred. No. 6.5e-163;
  Matches  667;  Conservative    0;  Mismatches    5;  Indels   21;  Gaps    3;

Qy      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    651 GTGATAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 710

Qy     61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    711 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 770

Qy    121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    771 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 830

Qy    181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    831 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 890

Qy    241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    891 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 950

Qy    301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    951 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 1010

Qy    361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1011 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 1070

Qy    421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1071 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 1130

Qy    481 GGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTC 540
          ||||||||||||||||||||          ||||||||||||||||||||||||||||||
Db   1131 GGGGCTCTGCCAGATCCTGAGA----------CGCTTCCCCTCCCTGCCCCACCCGGGTC 1180

Qy    541 CTGTGCTGGCTCCTGCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTG 590
          ||||||||||||||||||||          ||||||||||||||||||||||||||||||
Db   1181 CTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTG 1240

Qy    591 TGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGC 650
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   1241 TGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGA-CCCCTTGGGC 1299

Qy    651 TGAGCACCAAGACCTTGAACCTTTTTTGTTTTA 683
          |||||||||||||||||| |  ||| |||||
Db   1300 TGAGCACCAAGACCTTGAACATCATTTTTTTTA 1332


RESULT   9
AR144151
LOCUS       AR144151                669 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 16 from patent US 6210883.
ACCESSION   AR144151
VERSION     AR144151.1  GI:15106018
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   Reed,S.G. and Wang,T.Tong.
  TITLE     Compounds and methods for diagnosis of lung cancer
  JOURNAL   Patent: US 6210883-A 16 03-APR-2001;
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="unknown"
BASE COUNT      158 a    208 c    173 g    119 t     11 others
ORIGIN


  Query Match           41.9%;  Score 501.8;  DB 6;  Length 669;
  Best Local Similarity   90.5%;  Pred. No. 2.2e-131;
  Matches  616;  Conservative    0;  Mismatches   37;  Indels   28;  Gaps    7;

Qy     10 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 69
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 61

Qy     70 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 121

Qy    130 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 181
```

```
Qy    190 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 241

Qy    250 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 309
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    242 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCNGG 301

Qy    310 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 369
          |||||||||||||||||||||||| |||||||||||||||||||||||||||| |||||||
Db    302 AGACTACAACATGGCCAAAGCCAACATGAAGAATAAGCAGCTGCCAAGTGCANGACCAGA 361

Qy    370 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCCAGAGAAAGTC 429
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    362 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTG-CCCAGAGAAAGTC 420

Qy    430 CTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTG 489
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    421 CTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTG-CCGGGGCTCTG 479

Qy    490 CCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGG 549
          |||.|||            |       | |||||||||||||||||||||||||||||||
Db    480 CCANATC------------TGAGACGCTTCCCTCCCTGCCCCACCCGGGTCCTGTGCTGG 527

Qy    550 CTCCTGCCC---------CTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAG 600
          |||||||||         | ||||||  |||||||| |||||  | |||   |||||||
Db    528 CTCCTGCCCTTCCTGCTTTTGCAGCCANGGGTCAGGAAGTGGCNCNGGTNGTGGCTGGAA 587

Qy    601 AGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAA 660
          ||    || |||||||||||||||||||| | ||||||||||| ||||      |||
Db    588 AG--CAAAACCCTTTCCTGTTGGTGTCCCA-CCCATGGAGCCCCTGGGGC--GAGCCCAN 642

Qy    661 GACCTTGAACCTTTTTTGTTT 681
          || ||||| |||||||  | |
Db    643 GAACTTGANCCTTTTTGTTNT 663
```

```
RESULT  10
AR176338
LOCUS       AR176338                669 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 16 from patent US 6312695.
ACCESSION   AR176338
VERSION     AR176338.1  GI:17918693
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   Reed,S.G. and Wang,T.Tong.
  TITLE     Compounds and methods for therapy of lung cancer
  JOURNAL   Patent: US 6312695-A 16 06-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="unknown"
BASE COUNT      158 a    208 c    173 g    119 t     11 others
ORIGIN
```

```
  Query Match             41.9%;  Score 501.8;  DB 6;  Length 669;
  Best Local Similarity   90.5%;  Pred. No. 2.2e-131;
  Matches  616;  Conservative    0;  Mismatches   37;  Indels   28;  Gaps    7;

Qy     10 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 69
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 61

Qy     70 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 121

Qy    130 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    122 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 181

Qy    190 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 241

Qy    250 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 309
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    242 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCNGG 301

Qy    310 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 369
          |||||||||||||||||||||||| |||||||||||||||||||||||||||| |||||||
Db    302 AGACTACAACATGGCCAAAGCCAACATGAAGAATAAGCAGCTGCCAAGTGCANGACCAGA 361

Qy    370 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 429
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    362 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTG-CCCAGAGAAAGTC 420

Qy    430 CTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTG 489
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    421 CTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTG-CCGGGGCTCTG 479

Qy    490 CCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGG 549
          |||.|||            |       | |||||||||||||||||||||||||||||||
Db    480 CCANATC------------TGAGACGCTTCCCTCCCTGCCCCACCCGGGTCCTGTGCTGG 527

Qy    550 CTCCTGCCC---------CTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAG 600
          |||||||||         | ||||||  |||||||| |||||  | |||   |||||||
Db    528 CTCCTGCCCTTCCTGCTTTTGCAGCCANGGGTCAGGAAGTGGCNCNGGTNGTGGCTGGAA 587

Qy    601 AGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAA 660
          ||    || |||||||||||||||||||| | ||||||||||| ||||      |||
Db    588 AG--CAAAACCCTTTCCTGTTGGTGTCCCA-CCCATGGAGCCCCTGGGGC--GAGCCCAN 642

Qy    661 GACCTTGAACCTTTTTTGTTT 681
          || ||||| |||||||  | |
Db    643 GAACTTGANCCTTTTTGTTNT 663
```

```
RESULT  11
AF157510
LOCUS       AF157510                1093 bp    mRNA    linear   PRI 01-JUL-2000
DEFINITION  Homo sapiens beta endosulfine mRNA, complete cds.
ACCESSION   AF157510
VERSION     AF157510.1  GI:8886010
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1093)
  AUTHORS   Scott,V.E.S., Roch,J.-M., Davis-Taber,R.A., Molinari,E.J.,
            Whiteaker,K.L., Gopalakrishnan,M., Idler,K. and Sullivan,J.P.
  TITLE     Cloning and molecular characterization of two isoforms of human
            endosulfine
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1093)
  AUTHORS   Scott,V.E.S., Roch,J.-M., Whiteaker,K.L. and Sullivan,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-1999) D47C, Abbott Laboratories, AP10 L08, 100
            Abbott Park, Abbott Park, IL 60064, USA
FEATURES             Location/Qualifiers
     source          1. .1093
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             107. .460
                     /codon_start=1
                     /product="beta endosulfine"
                     /protein_id="AAF80341.1"
                     /db_xref="GI:8886011"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAG"
```

```
BASE COUNT      313 a    261 c    260 g    259 t
ORIGIN


  Query Match             38.0%;  Score 456;  DB 9;  Length 1093;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-118;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
           |||||||||||||||||||||||||||||||||||
Db     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456


RESULT  12
CNS01RIF/c
LOCUS       CNS01RIF              148578 bp    DNA     linear   PRI 26-APR-2001
DEFINITION  Human chromosome 14 DNA sequence BAC R-406A9 of library RPCI-11
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL163171
VERSION     AL163171.4  GI:11611169
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148578)
  AUTHORS   Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
  TITLE     Sequencing of the human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148578)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Dec 9, 2000 this sequence version replaced gi:9453784.
            ---------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            ----------
            The following BAC sequence is oriented from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : C-2007P7
            Downstream BAC (overlapping the SP6 end) : R-11K13 (AC=AL355095)
            ---------- Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 7.83x in Q20 bases; sum-of-contigs
            ----------
            ------------------------
            Overall quality chart :
            Range    :   bases
            0        :
            1  -  9  :       5
            10 - 19  :      94
            20 - 29  :     158
            30 - 39  :     446
            40 - 49  :    1989
            50 - 59  :    3410
            60 - 69  :    6039
            70 - 79  :   17886
            80 - 89  :   47912
            90 - 99  :   70639
            ------------------------
            Percentage of bases with a quality value >= 40 : 99 %.
FEATURES             Location/Qualifiers
     source          1. .148578
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="R-406A9"
                     /clone_lib="RPCI-11"
     STS             56891. .57046
                     /note="matching EMBL:R48962
                     RHdb:RH44475
                     dbSTS:STS37544
                     Identified using the e-PCR software (G. Schuler)"
     STS             139692. .139819
                     /note="matching EMBL:G04746
                     RHdb:RH53610
                     RHdb:RH3734
                     dbSTS:STS25994
                     Identified using the e-PCR software (G. Schuler)"
BASE COUNT    46908 a  26225 c  26858 g  48587 t
ORIGIN


  Query Match             36.2%;  Score 434.4;  DB 9;  Length 148578;
  Best Local Similarity   98.5%;  Pred. No. 9.6e-112;
  Matches  449;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           |||| ||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   59279 GTGATAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCTCTCCCCCTTCCCTTCTC 59220

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   59219 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAGCA 59160

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           |||||||||||||||||| |||||| ||||||||||||||||||||||||||||||||||
Db   59159 AGAAGAAGAGAACCCTGCAGAGGAGGCCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 59100

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   59099 AGGTATTCTCCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 59040

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   59039 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 58980

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   58979 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 58920
```

```
Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
           ||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   58919 AGGACCAGACAAGAACCTGGTGACTGGTGATCACAT-CCCACCCCACAGGATCTGCCCCA 58861

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
           ||||||||||||||||||||||||||||||||||
Db   58860 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 58825


RESULT  13
AC069588/c
LOCUS       AC069588              159846 bp    DNA     linear   HTG 04-OCT-2000
DEFINITION  Homo sapiens chromosome 14 clone RP11-340E1 map 14, WORKING DRAFT
            SEQUENCE, 22 unordered pieces.
ACCESSION   AC069588
VERSION     AC069588.3  GI:10567972
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159846)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 14, clone RP11-340E1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 159846)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 4, 2000 this sequence version replaced gi:8569195.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L8786
                Center clone name: 340_E_1
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 150080 bases at least Q40
                Consensus quality: 154891 bases at least Q30
                Consensus quality: 156760 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 157746; sum-of-contigs
                Quality coverage: 4.3 in Q20 bases; agarose-fp
                Quality coverage: 4.3 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    11674: contig of 11674 bp in length
            *   11675 11774: gap of      100 bp
            *   11775    12783: contig of 1009 bp in length
            *   12784 12883: gap of      100 bp
            *   12884    14380: contig of 1497 bp in length
            *   14381 14480: gap of      100 bp
            *   14481    15987: contig of 1507 bp in length
            *   15988 16087: gap of      100 bp
            *   16088    20096: contig of 4009 bp in length
            *   20097 20196: gap of      100 bp
            *   20197    24318: contig of 4122 bp in length
            *   24319 24418: gap of      100 bp
            *   24419    28596: contig of 4178 bp in length
            *   28597 28696: gap of      100 bp
            *   28697    32856: contig of 4160 bp in length
            *   32857 32956: gap of      100 bp
            *   32957    37590: contig of 4634 bp in length
            *   37591 37690: gap of      100 bp
            *   37691    43359: contig of 5669 bp in length
            *   43360 43459: gap of      100 bp
            *   43460    49215: contig of 5756 bp in length
            *   49216 49315: gap of      100 bp
            *   49316    55482: contig of 6167 bp in length
            *   55483 55582: gap of      100 bp
            *   55583    61888: contig of 6306 bp in length
            *   61889 61988: gap of      100 bp
            *   61989    70075: contig of 8087 bp in length
            *   70076 70175: gap of      100 bp
            *   70176    79049: contig of 8874 bp in length
            *   79050 79149: gap of      100 bp
            *   79150    87232: contig of 8083 bp in length
            *   87233 87332: gap of      100 bp
            *   87333    98465: contig of 11133 bp in length
            *   98466 98565: gap of      100 bp
            *   98566   116653: contig of 18088 bp in length
            *  116654 116753: gap of      100 bp
            *  116754   127734: contig of 10981 bp in length
            *  127735 127834: gap of      100 bp
            *  127835   139387: contig of 11553 bp in length
            *  139388 139487: gap of      100 bp
            *  139488   150540: contig of 11053 bp in length
            *  150541 150640: gap of      100 bp
            *  150641   159846: contig of 9206 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159846
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14"
                     /clone="RP11-340E1"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .11674
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    11775. .12783
                     /note="assembly_fragment"
     misc_feature    12884. .14380
                     /note="assembly_fragment"
     misc_feature    14481. .15987
                     /note="assembly_fragment"
     misc_feature    16088. .20096
                     /note="assembly_fragment"
     misc_feature    20197. .24318
```

```
                     /note="assembly_fragment"
     misc_feature    24419. .28596
                     /note="assembly_fragment"
     misc_feature    28697. .32856
                     /note="assembly_fragment"
     misc_feature    32957. .37590
                     /note="assembly_fragment"
     misc_feature    37691. .43359
                     /note="assembly_fragment"
     misc_feature    43460. .49215
                     /note="assembly_fragment"
     misc_feature    49316. .55482
                     /note="assembly_fragment"
     misc_feature    55583. .61888
                     /note="assembly_fragment"
     misc_feature    61989. .70075
                     /note="assembly_fragment"
     misc_feature    70176. .79049
                     /note="assembly_fragment"
     misc_feature    79150. .87232
                     /note="assembly_fragment"
     misc_feature    87333. .98465
                     /note="assembly_fragment"
     misc_feature    98566. .116653
                     /note="assembly_fragment"
     misc_feature    116754. .127734
                     /note="assembly_fragment"
     misc_feature    127835. .139387
                     /note="assembly_fragment"
     misc_feature    139488. .150540
                     /note="assembly_fragment"
     misc_feature    150641. .159846
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    49405 a  27703 c  27568 g  53067 t   2103 others
ORIGIN


  Query Match             36.2%;  Score 434.4;  DB 2;  Length 159846;
  Best Local Similarity   98.5%;  Pred. No.. 9.7e-112;
  Matches  449;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          |||| ||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db 119501 GTGATAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCTCTCCCCCTTCCCTTCTC 119442

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db 119441 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAGCA 119382

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          |||||||||||||||||| |||||| ||||||||||||||||||||||||||||||||||
Db 119381 AGAAGAAGAGAACCCTGCAGAGGAGGCCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 119322

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 119321 AGGTATTCTCCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 119262

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 119261 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 119202

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 119201 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 119142

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db 119141 AGGACCAGACAAGAACCTGGTGACTGGTGATCACAT-CCCACCCCACAGGATCTGCCCCA 119083

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
          |||||||||||||||||||||||||||||||||||
Db 119082 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 119047


RESULT  14
AC007449
LOCUS       AC007449              176616 bp    DNA     linear   HTG 13-APR-2000
DEFINITION  Homo sapiens clone RP11-104A1, WORKING DRAFT SEQUENCE, 13 unordered
            pieces.
ACCESSION   AC007449
VERSION     AC007449.2  GI:7547180
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176616)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-104A1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176616)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 13, 2000 this sequence version replaced gi:4726138.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L773
                Center clone name: 104_A_1
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 91% of reads
                Sequencing vector: Plasmid; n/a; %-0.f%% of reads
                8.87005649717514Chemistry: Dye-primer-amersham; 97% of reads
                Chemistry: Dye-terminator Big Dye; 3% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 170592 bases at least Q40
                Consensus quality: 173409 bases at least Q30
                Consensus quality: 174613 bases at least Q20
                Insert size: 180000; agarose-fp
                Insert size: 175416; sum-of-contigs
                Quality coverage.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1       51: contig of 51 bp in length
            *       52 151: gap of      100 bp
            *      152     1681: contig of 1530 bp in length
```

```
            *     1682 1781: gap of      100 bp
            *     1782     5071: contig of 3290 bp in length
            *     5072 5171: gap of      100 bp
            *     5172     9359: contig of 4188 bp in length
            *     9360 9459: gap of      100 bp
            *     9460    13792: contig of 4333 bp in length
            *    13793 13892: gap of      100 bp
            *    13893    23757: contig of 9865 bp in length
            *    23758 23857: gap of      100 bp
            *    23858    38876: contig of 15019 bp in length
            *    38877 38976: gap of      100 bp
            *    38977    52701: contig of 13725 bp in length
            *    52702 52801: gap of      100 bp
            *    52802    73312: contig of 20511 bp in length
            *    73313 73412: gap of      100 bp
            *    73413    92674: contig of 19262 bp in length
            *    92675 92774: gap of      100 bp
            *    92775   114928: contig of 22154 bp in length
            *   114929 115028: gap of      100 bp
            *   115029   143307: contig of 28279 bp in length
            *   143308 143407: gap of      100 bp
            *   143408   176616: contig of 33209 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176616
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-104A1"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .51
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    152. .1681
                     /note="assembly_fragment"
     misc_feature    1782. .5071
                     /note="assembly_fragment"
     misc_feature    5172. .9359
                     /note="assembly_fragment"
     misc_feature    9460. .13792
                     /note="assembly_fragment"
     misc_feature    13893. .23757
                     /note="assembly_fragment"
     misc_feature    23858. .38876
                     /note="assembly_fragment"
     misc_feature    38977. .52701
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    52802. .73312
                     /note="assembly_fragment"
     misc_feature    73413. .92674
                     /note="assembly_fragment"
     misc_feature    92775. .114928
                     /note="assembly_fragment"
     misc_feature    115029. .143307
                     /note="assembly_fragment"
     misc_feature    143408. .176616
                     /note="assembly_fragment"
BASE COUNT    52499 a  35498 c  34632 g  52787 t   1200 others
ORIGIN


  Query Match             32.0%;  Score 383.6;  DB 2;  Length 176616;
  Best Local Similarity   79.2%;  Pred. No. 2.7e-97;
  Matches  523;  Conservative    0;  Mismatches  119;  Indels   18;  Gaps    5;

Qy     538 GTCCTGTGCTGGCTCCTGCCCCTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTG 597
           | | | |||     ||  |   | ||||||||||||||||||||||||| |||  |||| |
Db   61106 GGCTTCTGCCCCTCCCATCTTTTGCAGCCAGGGGTCAGGAGGTGGCTTGGGCATGGGTTA 61165

Qy     598 GAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCAC 657
           ||||||||||||||| | ||  |  |||||||||||||||||||||||||||||||||||||
Db   61166 GAGAGGCAGAAGCCCCTGCCCATCAGTGTCCCAGCACATGGAGCCCCTTGGGCTAAGCAC 61225

Qy     658 CAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCA 717
           |||||||||||||| | ||||||||||||||||||||||||||||| ||||||||||| |||
Db   61226 CAAGACCTTGAACATGTTTTGTTTTACCTTTTTTCCAAATAACTGTTGGGAGAAATCTCA 61285

Qy     718 ATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGGAGGAGTAT 777
           |||||||  ||||||  |||  ||||       |||||||||||  ||||  |||||||  |||
Db   61286 ATGAAATCTTGGGGGCAGGGTTGGG------AGGGTGGGGTGAGAGACTTGGAGGAATAT 61339

Qy     778 GAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACGTGGCTGA 837
           |||||||  ||||||||||| ||||||||||||||||  |||||||||||||  | |||||
Db   61340 GAATTAGAGCTTGGAGTTGATAAAAACATTCCTGACGGTCCTTCTTAACCATGAGGCTGG 61399

Qy     838 TGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTAGTTGA 897
           ||||||  ||  ||||||    |||  |  | | ||  | | |  | | | | |||||||  |
Db   61400 TGTGGGATAAATATGAG----AGGGGGAAGTGGAGTCCACTAGAGATGAGATCTAGTTCA 61455

Qy     898 GCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATAAAAGAAA 957
           |||  ||  ||||||  ||||||||||||  |||||  ||||  ||||||||  ||||  |||||||
Db   61456 GCTTTGCAGATAACTGCCTTGTTTCCATGTGGCTGGAGGCCTGGTGTGAGATTAAAGAAA 61515

Qy     958 CTCCATCCGCACAGACAGATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAGTTGAGAACT 1017
           |||  |||  ||| ||||||||||||||||||||||||  |||||||   ||  |||  |||||||
Db   61516 CTCAATCTGCATAGACAGATGCAAACAGCTCCTCTGGTTCTGC--AGTTAGCTGAGAACA 61573

Qy    1018 CAACATTAATCA-----TTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGGGAAGA 1072
           |||  |||  | |      ||  |||   |||||||| ||||||||||| |||| || ||
Db   61574 CAATGTTAGTTATTCATTTAAAAAGCATACTGTCCTCGAAATAGATTTTCTGTTGGGAGG 61633

Qy    1073 AGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGG 1132
           ||||||||||||   || |||||| || |||| |||||| |||| ||||||||||||
Db   61634 AGGGCAGTGAGTTGTGGGGAAAGGAACCATGAGTGTGGGGAGCCCCA-AGAGCCCAAAGG 61692

Qy    1133 ACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAA 1192
           |||||||||||||| |||||||||   || ||||   |  ||||||   ||||| |  ||
Db   61693 ACTTTTTCAGTATTTGAAATAAACCCCACCAAAAAAAAAAAAAAAAAAAAAAAAAGACTAA 61752


RESULT  15
AC007124/c
LOCUS       AC007124               89263 bp    DNA     linear   PRI 20-MAR-1999
DEFINITION  Human genomic clone P1-11874, complete sequence.
ACCESSION   AC007124
VERSION     AC007124.1  GI:4454586
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 89263)
  AUTHORS   Nascimento,L., Spiegel,L.A., de la Bastide,M., Habermann,K.,
            See,L.H., Rodriguez,M., O'Shaughnessy,A., Matero,A., Swaby,I.,
            Huang,E.N., Schutz,K., Preston,R.R., Shah,R., Vil,M.D., Shekher,M.,
            Parnell,L.D., Dedhia,N.N. and McCombie,W.R.
  TITLE     Homo sapiens genomic clone P1-11874
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 89263)
  AUTHORS   Nascimento,L., Spiegel,L.A., de la Bastide,M., Habermann,K.,
            See,L.H., Rodriguez,M., O'Shaughnessy,A., Matero,A., Swaby,I.,
            Huang,E.N., Schutz,K., Preston,R.R., Shah,R., Vil,M.D., Shekher,M.,
            Parnell,L.D., Dedhia,N.N. and McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-1999) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724
COMMENT     Position 1 of P1-11874 is at position 16011 of cloning vector
            pSacBII and position 89263 is at position 1 of pSacBII. P1-11874 is
            40.6% G+C.
FEATURES             Location/Qualifiers
     source          1. .89263
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT    25953 a  18074 c  18201 g  27035 t
```

```
ORIGIN

  Query Match             31.8%;  Score 381.6;  DB 9;  Length 89263;
  Best Local Similarity   79.2%;  Pred. No. 8.9e-97;
  Matches  521;  Conservative    0;  Mismatches  119;  Indels   18;  Gaps    5;

Qy     538 GTCCTGTGCTGGCTCCTGCCCCTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTG 597
           | | | |||     ||   |   | ||||||||||||||||||||||| |||  |||| |
Db   43215 GGCTTCTGCCCCTCCCATCTTTTGCAGCCAGGGGTCAGGAGGTGGCTTGGGCATGGGTTA 43156

Qy     598 GAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCAC 657
           |||||||||||||||| | ||  |  ||||||||||||||||||||||||||||||| |||||
Db   43155 GAGAGGCAGAAGCCCCTGCCCATCAGTGTCCCAGCACATGGAGCCCCTTGGGCTAAGCAC 43096

Qy     658 CAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCA 717
           |||||||||||| | ||||||||||||||||||||||||| |||||||||||| |||
Db   43095 CAAGACCTTGAACATGTTTTGTTTTACCTTTTTTCCAAATAACTGTTGGGAGAAATCTCA 43036

Qy     718 ATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGGAGGAGTAT 777
           |||||||  |||||||  ||| ||||      ||||||||||| ||||  |||||||  |||
Db   43035 ATGAAATCTTGGGGGTAGGGTTGGG------AGGGTGGGGTGAGAGACTTGGAGGAATAT 42982

Qy     778 GAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACGTGGCTGA 837
           |||||||  ||||||||||| ||||||||||||||||  |||||||||||||.| |||||
Db   42981 GAATTAGAGCTTGGAGTTGATAAAAACATTCCTGACGGTCCTTCTTAACCATGAGGCTGG 42922

Qy     838 TGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTAGTTGA 897
           ||||||  ||  ||||||    |||  |  | | || | | |  | | | | |||||||  |
Db   42921 TGTGGGATAAATATGAG----AGGGGGAAGTGGAGTCCACTAGAGATGAGATCTAGTTCA 42866

Qy     898 GCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATAAAAGAAA 957
           |||  || ||||||  ||||||||||  |||||  |||||  ||||||||  ||||  |||||||
Db   42865 GCTTTGCAGATAACTGCCTTGTTTCCATGTGGCTGGAGGCCTGGTGTGAGATTAAAGAAA 42806

Qy     958 CTCCATCCGCACAGACAGATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAGTTGAGAACT 1017
           ||  |||  |||  ||||||||||||||||||||||||  |||||||  ||  |||  |||||||
Db   42805 CTGAATCTGCATAGACAGATGCAAACAGCTCCTCTGGTTCTGC--AGTTAGCTGAGAACA 42748

Qy    1018 CAACATTAATCA-----TTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGGGAAGA 1072
           |||  |||  | |     ||  |||   |||||||||  ||||||||||  |||| || ||
Db   42747 CAATGTTAGTTATTCATTTAAAAAGCATACTGTCCTCGAAATAGATTTTCTGTTGGGAGG 42688

Qy    1073 AGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGG 1132
           ||||||||||||   || |||||||  || ||||  |||||||  |||||  ||||||||||||
Db   42687 AGGGCAGTGAGTTGTGGGGAAAGGAACCATGAGTGTGGGGAGCCCCA-AGAGCCCAAAGG 42629

Qy    1133 ACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAA 1190
           ||||||||||||||  |||||||||   || ||||   |  ||||||   |||  |  ||
Db   42628 ACTTTTTCAGTATTTGAAATAAACCCCACCAAAAAAAAAAAAAAAAAAAAAAAGACTAA 42571


RESULT  16
HSJ822J19/c
LOCUS       HSJ822J19              89872 bp    DNA     linear   PRI 19-MAR-2001
DEFINITION  Human DNA sequence from clone RP5-822J19 on chromosome 20. Contains
            an alpha-endosulfine pseudogene, STSs and GSSs, complete sequence.
ACCESSION   AL050324
VERSION     AL050324.5  GI:5002694
KEYWORDS    HTG; endosulfine.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 89872)
  AUTHORS   Matthews,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 7, 1999 this sequence version replaced gi:4972283.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            RP5-822J19 is from the library RPCI-5 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP5-822J19.
FEATURES             Location/Qualifiers
     source          1. .89872
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP5-822J19"
                     /clone_lib="RPCI-5"
     repeat_region   1. .112
                     /note="MER5B repeat: matches 67. .178 of consensus"
     misc_feature    113. .601
                     /note="match: STS: Em:HS822J19S"
     repeat_region   833. .865
                     /note="MER5A repeat: matches 14. .46 of consensus"
     repeat_region   1006. .1313
                     /note="AluSg repeat: matches 1. .308 of consensus"
     repeat_region   1349. .1679
                     /note="LTR16C repeat: matches 53. .377 of consensus"
     repeat_region   1979. .2170
                     /note="MER5A repeat: matches 1. .186 of consensus"
     repeat_region   2526. .2859
                     /note="L1M4 repeat: matches 4254. .4604 of consensus"
     repeat_region   2862. .3211
                     /note="THE1A repeat: matches 1. .354 of consensus"
     repeat_region   3212. .4798
                     /note="THE1A-internal repeat: matches 1. .1580 of
                     consensus"
     repeat_region   4799. .5152
                     /note="THE1A repeat: matches 1. .354 of consensus"
     repeat_region   5194. .5680
                     /note="L1MC1 repeat: matches 5813. .6317 of consensus"
     repeat_region   6026. .6232
                     /note="MIR repeat: matches 35. .247 of consensus"
     repeat_region   6419. .6715
                     /note="AluJb repeat: matches 1. .309 of consensus"
     repeat_region   6727. .6850
                     /note="31 copies 4 mer aaga 76% conserved"
     repeat_region   6737. .6850
                     /note="57 copies 2 mer ga 75% conserved"
     repeat_region   7405. .7434
                     /note="15 copies 2 mer aa 100% conserved"
     repeat_region   7859. .8091
                     /note="MIR repeat: matches 16. .249 of consensus"
     misc_feature    complement(8145. .8666)
                     /note="match: GSS: Em:AQ603396"
     misc_feature    complement(8339. .8673)
                     /note="match: STS: Em:HSPF05H8"
     misc_feature    complement(9561. .10143)
                     /note="match: GSS: Em:AQ582705"
```

```
repeat_region   10445. .10777
                /note="L1 repeat: matches 4859. .5202 of consensus"
repeat_region   10928. .11110
                /note="L1MB3 repeat: matches 5981. .6181 of consensus"
repeat_region   11722. .11864
                /note="MIR repeat: matches 112. .260 of consensus"
repeat_region   13615. .13909
                /note="AluSx repeat: matches 1. .294 of consensus"
repeat_region   14917. .15218
                /note="AluSc repeat: matches 1. .289 of consensus"
repeat_region   15341. .15443
                /note="MIR repeat: matches 72. .197 of consensus"
repeat_region   17565. .17859
                /note="AluY repeat: matches 1. .295 of consensus"
repeat_region   17994. .18333
                /note="LTR16A repeat: matches 96. .447 of consensus"
repeat_region   19036. .19340
                /note="AluSx repeat: matches 1. .300 of consensus"
repeat_region   20208. .20306
                /note="MER5A repeat: matches 86. .187 of consensus"
repeat_region   21264. .21315
                /note="13 copies 4 mer gtgt 86% conserved"
repeat_region   21265. .21314
                /note="25 copies 2 mer tg 88% conserved"
repeat_region   21844. .22314
                /note="L2 repeat: matches 1997. .2502 of consensus"
repeat_region   22539. .22839
                /note="AluSp repeat: matches 1. .302 of consensus"
repeat_region   22854. .22984
                /note="FLAM_C repeat: matches 3. .133 of consensus"
repeat_region   22985. .23018
                /note="BC200 repeat: matches 1. .161 of consensus"
misc_feature    complement(23059. .23343)
                /note="match: STS: Em:G07642"
repeat_region   23344. .23651
                /note="AluSx repeat: matches 1. .308 of consensus"
misc_feature    complement(23652. .23753)
                /note="match: GSS: Em:AQ807387"
misc_feature    complement(23652. .23754)
                /note="match: GSS: Em:AQ544217"
misc_feature    complement(23652. .23752)
                /note="match: GSS: Em:AQ544233"
misc_feature    23754. .24328
                /note="match: GSS: Em:AQ346979"
misc_feature    23900. .24386
                /note="match: GSS: Em:AQ711642"
repeat_region   24606. .24731
                /note="63 copies 2 mer tt 59% conserved"
misc_feature    complement(24848. .25354)
                /note="match: GSS: Em:AQ561985"
misc_feature    complement(24849. .25354)
                /note="match: GSS: Em:AQ561966"
misc_feature    25355. .25824
                /note="match: GSS: Em:AQ345431"
misc_feature    25358. .26074
                /note="match: GSS: Em:AQ744401"
misc_feature    25369. .25875
                /note="match: GSS: Em:AQ693363"
repeat_region   26075. .26496
                /note="MLT1C repeat: matches 1. .438 of consensus"
repeat_region   26808. .26833
                /note="13 copies 2 mer tt 96% conserved"
repeat_region   26953. .27167
                /note="MIR repeat: matches 36. .250 of consensus"
misc_feature    complement(29003. .29207)
                /note="match: GSS: Em:AQ101496"
repeat_region   29992. .30051
                /note="MIR repeat: matches 195. .254 of consensus"
misc_feature    31069. .31571
                /note="match: GSS: Em:AQ371017"
misc_feature    31071. .31481
                /note="match: GSS: Em:AQ424527"
misc_feature    31073. .31596
                /note="match: GSS: Em:AQ568767"
repeat_region   31914. .32221
                /note="AluSx repeat: matches 1. .307 of consensus"
repeat_region   33461. .33764
                /note="L2 repeat: matches 2379. .2750 of consensus"
repeat_region   33912. .33935
                /note="12 copies 2 mer tt 95% conserved"
repeat_region   34564. .35249
                /note="LTR8 repeat: matches 1. .691 of consensus"
gene            complement(35297. .35637)
                /gene="dJ822J19.1"
CDS             complement(35297. .35637)
                /gene="dJ822J19.1"
                /note="dJ822J19.1 (alpha-endosulfine pseudogene)
                match: proteins: Tr:Q9Z0G1 Tr:O97976 Sw:P56212 Sw:Q28055
                Tr:P79322 Tr:O43768"
                /codon_start=1
                /pseudo
                /evidence=not_experimental
repeat_region   35829. .35885
                /note="L2 repeat: matches 2036. .2093 of consensus"
misc_feature    36012. .36251
                /note="match: GSS: Em:AQ243178"
repeat_region   36252. .36402
                /note="MIR repeat: matches 99. .262 of consensus"
repeat_region   36590. .36617
                /note="7 copies 4 mer tgtt 96% conserved"
repeat_region   38149. .38220
                /note="MER44 repeat: matches 657. .728 of consensus"
repeat_region   38229. .38490
                /note="MER44 repeat: matches 396. .653 of consensus"
repeat_region   38548. .38816
                /note="MER44A repeat: matches 1. .266 of consensus"
repeat_region   39491. .39566
                /note="19 copies 4 mer tatc 90% conserved"

 Query Match            31.7%;  Score 380;  DB 9;  Length 89872;
 Best Local Similarity  79.0%;  Pred. No. 2.5e-96;
 Matches  520;  Conservative    0;  Mismatches  120;  Indels   18;  Gaps    5;

Qy    538 GTCCTGTGCTGGCTCCTGCCCCTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTG 597
          | | | |||     ||  |   | ||||||||||||||||||||||| |||  |||| |
Db  34552 GGCTTCTGCCCCTCCCATCTTTTGCAGCCAGGGGTCAGGAGGTGGCTTGGGCATGGGTTA 34493

Qy    598 GAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCAC 657
          |||||||||||||||| | ||  |  |||||||||||||||||||||||||||||| |||||
Db  34492 GAGAGGCAGAAGCCCCTGCCCATCAGTGTCCCAGCACATGGAGCCCCTTGGGCTAAGCAC 34433

Qy    658 CAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCA 717
          |||||||||||||  | ||||||||||||||||||||||||||| ||||||||||| |||
Db  34432 CAAGACCTTGAACATGTTTTGTTTTACCTTTTTTCCAAATAACTGTTGGGAGAAATCTCA 34373

Qy    718 ATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGGAGGAGTAT 777
          |||||||  ||||||  ||| ||||      ||||||||||| ||||  |||||| |||
Db  34372 ATGAAATCTTGGGGGCAGGGTTGGG------AGGGTGGGGTGAGAGACTTGGAGGAATAT 34319

Qy    778 GAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACGTGGCTGA 837
          |||||||  |||||||||||  ||||||||||||||||   ||||||||||||  | |||||
Db  34318 GAATTAGAGCTTGGAGTTGATAAAAACATTCCTGACGGTCCTTCTTAACCATGAGGCTGG 34259

Qy    838 TGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTAGTTGA 897
          ||||||  ||  ||||||    |||  |  | | ||  | | |   | | | | |||||||  |
Db  34258 TGTGGGATAAATATGAG----AGGGGGAAGTGGAGTCCACTAGAGATGAGATCTAGTTCA 34203

Qy    898 GCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATAAAAGAAA 957
          ||| ||  ||||||  ||||||||||  |||||  |||||  ||||||||  ||||  |||||||
Db  34202 GCTTTGCAGATAACTGCCTTGTTTCCATGTGGCTGGAGGCCTGGTGTGAGATTAAAGAAA 34143

Qy    958 CTCCATCCGCACAGACAGATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAGTTGAGAACT 1017
          |||  |||  |||  ||||||||||||||||||||||||  |||||||   ||  |||  |||||||
Db  34142 CTCAATCTGCATAGACAGATGCAAACAGCTCCTCTGGTTCTGC--AGTTAGCTGAGAACA 34085

Qy   1018 CAACATTAATCA-----TTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGGGAAGA 1072
```

```
          ||| ||| | |     ||  |||   ||||||||| |||||||||| |||| || ||
Db  34084 CAATGTTAGTTATTCATTTAAAAAGCATACTGTCCTCGAAATAGATTTTCTGTTGGGAGG 34025

Qy   1073 AGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGG 1132
          ||||||||||||   || |||||| || |||| |||||| ||||| ||||||||||||
Db  34024 AGGGCAGTGAGTTGTGGGGAAAGGAACCATGAGTGTGGGGAGCCCCA-AGAGCCCAAAGG 33966

Qy   1133 ACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAA 1190
          |||||||||||||| |||||| ||   || ||||   |  ||||||   ||| |  ||
Db  33965 ACTTTTTCAGTATTTGAAATACACCCCACCAAAAAAAAAAAAAAAAAAAAAAAGACTAA 33908


RESULT  17
SSA005987
LOCUS       SSA005987               366 bp    mRNA    linear   MAM 07-JAN-1999
DEFINITION  Sus scrofa alpha-endosulfine (complete coding sequence).
ACCESSION   AJ005987
VERSION     AJ005987.1  GI:4138556
KEYWORDS    alpha-endosulfine.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Bataille D., U 376 'Endocrinologie des
            Peptides et regulation genique', INSERM, CHU Arnaud-de-Villeneuve,
            34295, FRANCE
REFERENCE   2  (bases 1 to 366)
  AUTHORS   Virsolvy-Vergine,A., Salazar,G., Sillard,R., Denoroy,L., Mutt,V.
            and Bataille,D.
  TITLE     Endosulfine, endogenous ligand for the sulphonylurea receptor:
            isolation from porcine brain and partial structural determination
            of the alpha form
  JOURNAL   Diabetologia 39 (2), 135-141 (1996)
  MEDLINE   96218590
REFERENCE   3  (bases 1 to 366)
  AUTHORS   Peyrollier,K., Heron,L., Virsolvy-Vergine,A., Le Cam,A. and
            Bataille,D.
  TITLE     Alpha endosulfine is a novel molecule, structurally related to a
            family of phosphoproteins
  JOURNAL   Biochem. Biophys. Res. Commun. 223 (3), 583-586 (1996)
  MEDLINE   96268722
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /tissue_type="brain"
     CDS             1. .366
                     /function="endogenous regulator of K-ATP channels"
                     /codon_start=1
                     /evidence=experimental
                     /product="alpha-endosulfine"
                     /protein_id="CAA06801.1"
                     /db_xref="GI:4138557"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPEKAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAGGQVE"
BASE COUNT      129 a     91 c     96 g     50 t
ORIGIN


  Query Match           29.2%;  Score 350;  DB 4;  Length 366;
  Best Local Similarity   97.3%;  Pred. No. 2.8e-88;
  Matches  356;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy     107 ATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGAC 166
           |||||||||||||||||||||| ||||||||||||| ||||||||||||||||||||
Db       1 ATGTCCCAGAAACAAGAAGAAGAAAACCCTGCGGAGGAAACCGGCGAGGAGAAGCAGGAT 60

Qy     167 ACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAA 226
           || |||||||||||||||||||| ||||||| |||||| || ||||||||||||||||||
```

```
Db      61 ACACAGGAGAAAGAAGGTATTCTTCCTGAGAAAGCTGAGGAAGCAAAGCTAAAGGCCAAA 120

Qy     227 TACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAA 286
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TATCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAA 180

Qy     287 GGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAG 346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAG 240

Qy     347 CAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCA 406
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     241 CAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGACCACATCCCCACCCCA 300

Qy     407 CAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTT 466
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTT 360

Qy     467 GAATGA 472
           ||||||
Db     361 GAATGA 366


RESULT  18
BTA005986
LOCUS       BTA005986               366 bp    mRNA    linear   MAM 07-JAN-1999
DEFINITION  Bos taurus alpha-endosulfine (complete coding sequence).
ACCESSION   AJ005986
VERSION     AJ005986.1  GI:4127577
KEYWORDS    alpha-endosulfine.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Bataille D., U 376 'Endocrinologie des
            Peptides et regulation genique', INSERM, CHU Arnaud-de-Villeneuve,
            34295, FRANCE
REFERENCE   2  (bases 1 to 366)
  AUTHORS   Virsolvy-Vergine,A., Salazar,G., Sillard,R., Denoroy,L., Mutt,V.
            and Bataille,D.
  TITLE     Endosulfine, endogenous ligand for the sulphonylurea receptor:
            isolation from porcine brain and partial structural determination
            of the alpha form
  JOURNAL   Diabetologia 39 (2), 135-141 (1996)
  MEDLINE   96218590
REFERENCE   3  (bases 1 to 366)
  AUTHORS   Peyrollier,K., Heron,L., Virsolvy-Vergine,A., Le Cam,A. and
            Bataille,D.
  TITLE     Alpha endosulfine is a novel molecule, structurally related to a
            family of phosphoproteins
  JOURNAL   Biochem. Biophys. Res. Commun. 223 (3), 583-586 (1996)
  MEDLINE   96268722
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /tissue_type="brain"
     CDS             1. .366
                     /function="endogenous regulator of K-ATP channels"
                     /codon_start=1
                     /evidence=experimental
                     /product="alpha-endosulfine"
                     /protein_id="CAA06800.1"
                     /db_xref="GI:4127578"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPEKAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAGGQVE"
BASE COUNT      129 a     94 c     96 g     47 t
ORIGIN
```

```
  Query Match           29.1%;  Score 348.4;  DB 4;  Length 366;
  Best Local Similarity  97.0%;  Pred. No. 8.1e-88;
  Matches  355;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy     107 ATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGAC 166
           ||||||||||||||||||||||| |||||||||||||| ||||||||||||||||||||
Db       1 ATGTCCCAGAAACAAGAAGAAGAAAACCCTGCGGAGGAAACCGGCGAGGAGAAGCAGGAT 60

Qy     167 ACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAA 226
           || |||||||||||||||||| ||||||| |||||| || ||||||||||||||||||
Db      61 ACACAGGAGAAAGAAGGTATTCTTCCTGAGAAAGCTGAGGAAGCAAAGCTAAAGGCCAAA 120

Qy     227 TACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAA 286
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAA 180

Qy     287 GGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAG 346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAG 240

Qy     347 CAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCA 406
           |||||||||||||||||||||||||||||||||||| ||||| ||||||||||||||||
Db     241 CAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACCGGTGACCACATCCCCACCCCA 300

Qy     407 CAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTT 466
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CAGGACCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTT 360

Qy     467 GAATGA 472
           ||||||
Db     361 GAATGA 366


RESULT  19
MMA005985
LOCUS       MMA005985               366 bp    mRNA    linear   ROD 07-JAN-1999
DEFINITION  Mus musculus alpha-endosulfine (complete coding sequence).
ACCESSION   AJ005985
VERSION     AJ005985.1  GI:4138191
KEYWORDS    alpha-endosulfine.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Bataille D., U 376 'Endocrinologie des
            Peptides et regulation genique', INSERM, CHU Arnaud-de-Villeneuve,
            34295, FRANCE
REFERENCE   2  (bases 1 to 366)
  AUTHORS   Virsolvy-Vergine,A., Salazar,G., Sillard,R., Denoroy,L., Mutt,V.
            and Bataille,D.
  TITLE     Endosulfine, endogenous ligand for the sulphonylurea receptor:
            isolation from porcine brain and partial structural determination
            of the alpha form
  JOURNAL   Diabetologia 39 (2), 135-141 (1996)
  MEDLINE   96218590
REFERENCE   3  (bases 1 to 366)
  AUTHORS   Peyrollier,K., Heron,L., Virsolvy-Vergine,A., Le Cam,A. and
            Bataille,D.
  TITLE     Alpha endosulfine is a novel molecule, structurally related to a
            family of phosphoproteins
  JOURNAL   Biochem. Biophys. Res. Commun. 223 (3), 583-586 (1996)
  MEDLINE   96268722
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /tissue_type="brain"
     CDS             1. .366
                     /function="endogenous regulator of K-ATP channels"
                     /codon_start=1
                     /evidence=experimental
                     /product="alpha-endosulfine"
                     /protein_id="CAA06799.1"
                     /db_xref="GI:4138192"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPEKAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGADKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAGGQVE"
BASE COUNT      127 a     94 c    100 g     45 t
ORIGIN


  Query Match           29.1%;  Score 348.4;  DB 10;  Length 366;
  Best Local Similarity  97.0%;  Pred. No. 8.1e-88;
  Matches  355;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy     107 ATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGAC 166
           |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db       1 ATGTCCCAGAAACAAGAAGAAGAAAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGAT 60

Qy     167 ACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAA 226
           || |||||||||||||| ||||| ||||||| |||||| ||||||||||||||||||||
Db      61 ACACAGGAGAAAGAAGGGATTCTCCCTGAGAAAGCTGAGGAGGCAAAGCTAAAGGCCAAA 120

Qy     227 TACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAA 286
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAA 180

Qy     287 GGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAG 346
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     181 GGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAACAAG 240

Qy     347 CAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCA 406
           |||||||||||||||||| |||||||||||||||||| ||||| ||||||||||||||||
Db     241 CAGCTGCCAAGTGCAGGAGCAGACAAGAACCTGGTGACCGGTGACCACATCCCCACCCCA 300

Qy     407 CAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTT 466
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTT 360

Qy     467 GAATGA 472
           ||||||
Db     361 GAATGA 366


RESULT  20
RNA005984
LOCUS       RNA005984               366 bp    mRNA    linear   ROD 07-JAN-1999
DEFINITION  Rattus norvegicus mRNA alpha-endosulfine (complete coding
            sequence).
ACCESSION   AJ005984
VERSION     AJ005984.1  GI:4138425
KEYWORDS    Rattus norvegicus.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Bataille D., U 376 'Endocrinologie des
            Peptides et regulation genique', INSERM, CHU Arnaud-de-Villeneuve,
            34295, FRANCE
REFERENCE   2  (bases 1 to 366)
  AUTHORS   Virsolvy-Vergine,A., Salazar,G., Sillard,R., Denoroy,L., Mutt,V.
            and Bataille,D.
  TITLE     Endosulfine, endogenous ligand for the sulphonylurea receptor:
            isolation from porcine brain and partial structural determination
            of the alpha form
  JOURNAL   Diabetologia 39 (2), 135-141 (1996)
  MEDLINE   96218590
```

```
REFERENCE   3  (bases 1 to 366)
  AUTHORS   Peyrollier,K., Heron,L., Virsolvy-Vergine,A., Le Cam,A. and
            Bataille,D.
  TITLE     Alpha endosulfine is a novel molecule, structurally related to a
            family of phosphoproteins
  JOURNAL   Biochem. Biophys. Res. Commun. 223 (3), 583-586 (1996)
  MEDLINE   96268722
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /tissue_type="brain"
     CDS             1. .366
                     /function="endogenous regulator of K-ATP channels"
                     /note="Isolated as an endogenous ligand for the
                     sulphonylurea receptor"
                     /codon_start=1
                     /evidence=experimental
                     /product="alpha-endosulfine"
                     /protein_id="CAA06798.1"
                     /db_xref="GI:4138426"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPEKAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGADKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAGGQVE"
BASE COUNT      129 a     90 c     97 g     50 t
ORIGIN


  Query Match           28.8%;  Score 345.2;  DB 10;  Length 366;
  Best Local Similarity   96.4%;  Pred. No. 6.6e-87;
  Matches  353;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy    107 ATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGAC 166
          ||||||||||||||||||||||| ||||||||||||||| ||||||||||||||||||||
Db      1 ATGTCCCAGAAACAAGAAGAAGAAAACCCTGCGGAGGAAACCGGCGAGGAGAAGCAGGAT 60

Qy    167 ACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAA 226
          || |||||||||||||||||| ||||||| |||||| || ||||||||||||||||||||
Db     61 ACACAGGAGAAAGAAGGTATTCTTCCTGAGAAAGCTGAGGAAGCAAAGCTAAAGGCCAAA 120

Qy    227 TACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAA 286
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TATCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAA 180

Qy    287 GGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAG 346
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    181 GGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAACAAG 240

Qy    347 CAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCA 406
          ||||| || ||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    241 CAGCTACCGAGTGCAGGAGCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCA 300

Qy    407 CAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTT 466
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTT 360

Qy    467 GAATGA 472
          ||||||
Db    361 GAATGA 366


RESULT  21
AC016708
LOCUS       AC016708              194304 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens BAC clone RP11-223M6 from 2, complete sequence.
ACCESSION   AC016708
VERSION     AC016708.11  GI:18093082
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194304)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 194304)
  AUTHORS   Yoakum,M., Harkins,C. and Haakenson,W.
  TITLE     The sequence of Homo sapiens BAC clone RP11-223M6
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 194304)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 194304)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jan 9, 2002 this sequence version replaced gi:13431120.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0223M06
            --------------

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-172J1.  Actual start of
            this clone is at base position 1 of RP11-223M6; actual end is at
            base position 194304 of RP11-223M6.

            Data from AC068051 was used to finish the clone, AC016708.
            Polymorphisms have been identified between AC016708  and AC068051.
            An unresolved tandem repeat exists between 141687-142816.
            Approximately 600 bases are thought to be missing.
FEATURES             Location/Qualifiers
     source          1. .194304
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
```

```
                /clone="RP11-223M6"
                /clone_lib="RPCI-11"
repeat_region   173. .204
                /rpt_family="A-rich"
repeat_region   294. .473
                /rpt_family="MaLR"
repeat_region   555. .716
                /rpt_family="MaLR"
repeat_region   1112. .1414
                /rpt_family="MaLR"
repeat_region   2301. .2326
                /rpt_family="AT_rich"
repeat_region   2695. .4442
                /rpt_family="L1"
misc_feature    3084. .3086
                /note="similar to Homo sapiens EST AA765455 (NID:g2816693)
                oa03a11.s1"
misc_feature    4443. .4524
                /note="similar to Homo sapiens EST BI770278
                (NID:g15761856)"
misc_feature    4445. .5222
                /note="similar to Mus musculus EST BI156916
                (NID:g14616917)"
misc_feature    4445. .4524
                /note="similar to Homo sapiens EST BG538360
                (NID:g13530593)"
misc_feature    4449. .4875
                /note="similar to Mus musculus EST BE624006 (NID:g9904422)
                uu46g10.y1"
misc_feature    4450. .4524
                /note="similar to Homo sapiens EST AI929700 (NID:g5665664)
                au63f01.y1"
misc_feature    4454. .4524
                /note="similar to Homo sapiens EST BI463649
                (NID:g15254305)"
misc_feature    4468. .4524
                /note="similar to Homo sapiens EST BI255301
                (NID:g14808547)"
misc_feature    4474. .5306
                /note="similar to Mus musculus EST BI661673
                (NID:g15575909)"
misc_feature    4578. .4870
                /note="similar to Homo sapiens EST BI255301
                (NID:g14808547)"
misc_feature    4578. .5003
                /note="similar to Homo sapiens EST BI770278
                (NID:g15761856)"
misc_feature    4578. .4917
                /note="similar to Homo sapiens EST AI929700 (NID:g5665664)
                au63f01.y1"
misc_feature    4578. .5222
                /note="similar to Homo sapiens EST BI463649
                (NID:g15254305)"
misc_feature    4578. .4875
                /note="similar to Homo sapiens EST BG538360
                (NID:g13530593)"
misc_feature    4636. .4917
                /note="similar to Homo sapiens EST BE830880
                (NID:g10263258)"
misc_feature    4636. .4821
                /note="similar to Homo sapiens EST AI658639 (NID:g4762209)
                tu06g05.x1"
misc_feature    4637. .5179
                /note="similar to Bos taurus EST AV599105 (NID:g9717850)"
misc_feature    4672. .5003
                /note="similar to Homo sapiens EST AA349861
                (NID:g2002181)"
misc_feature    4685. .4875
                /note="similar to Homo sapiens EST AW392207
                (NID:g6896866)"
misc_feature    4744
                /note="similar to Homo sapiens EST BG202037
                (NID:g13723724)"
misc_feature    4755. .5515
                /note="similar to Homo sapiens EST BG202037
                (NID:g13723724)"
misc_feature    4758. .5152
                /note="similar to Mus musculus EST BI664062
                (NID:g15578295)"
misc_feature    4858. .5232
                /note="similar to Homo sapiens EST H70078 (NID:g1040284)
                yr92b08.r1"
misc_feature    4892. .5499
                /note="similar to Homo sapiens EST BF997045
                (NID:g12403368)"
misc_feature    4894. .5062
                /note="similar to Homo sapiens EST BF992134
                (NID:g12398457)"
misc_feature    4922. .5507
                /note="similar to Homo sapiens EST BF997056
                (NID:g12403379)"
misc_feature    4943. .5500
                /note="similar to Homo sapiens EST BI015134
                (NID:g14419205)"
misc_feature    4987. .5719
                /note="similar to Homo sapiens EST BI856660
                (NID:g15997407)"
misc_feature    5101. .5520
                /note="similar to Homo sapiens EST AA765455 (NID:g2816693)
                oa03a11.s1"
misc_feature    5120. .5232
                /note="similar to Homo sapiens EST BF992134
                (NID:g12398457)"
misc_feature    5257. .5643
                /note="similar to Homo sapiens EST BF725771
                (NID:g12041682) bx19g05.y1"
misc_feature    5265. .5269
                /note="similar to Homo sapiens EST BF305685
                (NID:g11252669)"
misc_feature    5265. .5269
                /note="similar to Homo sapiens EST BF305826
                (NID:g11252835)"
misc_feature    5307. .5759
                /note="similar to Homo sapiens EST AI360592 (NID:g4112213)
                qx64d07.x1"
misc_feature    5391. .5923
                /note="similar to Homo sapiens EST BF305826
```

Query Match 24.0%; Score 287.4; DB 9; Length 194304;
Best Local Similarity 81.8%; Pred. No. 6.3e-70;
Matches 374; Conservative 0; Mismatches 66; Indels 17; Gaps 3;

```
Qy      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          |||||||||||| ||||||||||  |||||||||| |||||| ||||| ||||||| ||
Db   4468 GTGACAGGAGCCAAAGCAGCAGCATAGGTTGTCCCCATTTCCCTTCCCCTTTCCCTTGTC 4527

Qy     61 CGGTTGCCTTCCCGGGCCC-CTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAAC 119
           ||| | ||    |||||| | ||  | ||| || | | |||| ||| |||||| |||
Db   4528 AGGTGGACTATATGGGCCCTCCTATGCCCCAGAGACTCAGTCCTACCACATCCCAGTAAC 4587

Qy    120 AAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAG 179
          |||||  |||||||||||| |||| |||||  | ||||||| | | | | |||||| ||||
Db   4588 AAGAAAGAGAGAACCCTGTGGAGAAGACCAACAAGGAGAAAC-GAAGATGCAGGATAAAG 4646

Qy    180 AAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAG 239
          ||||||||||  |||||||| ||| ||| |||||||||||| |||||||| ||||||||
Db   4647 AAGGTATTCTCTCTGAGAGAGTTGAGGAGCCAAAGCTAAAGTCCAAATACCTAAGCCTAG 4706

Qy    240 GACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACT 299
          ||||| |||| |||||||||| |||||||||||||||| | |||||||| ||| ||||||
Db   4707 GACAAGAGCCCGGAGGCTCCAACTTCCTCATGAAGAGATTACAGAAAGGGAAAAGGTACT 4766

Qy    300 TTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTG 359
          ||| |||||||||||||||||||||||||||||||||||||||| |||||||| |||||||
Db   4767 TTGTCTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAGTAAGCAGCTTCCAAGTG 4826
```

```
Qy     360 CAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCC 419
           ||| |||| | |||||||||||||||||||| |||||                ||||||||
Db    4827 CAGAACCAAAGAAGAACCTGGTGACTGGTGACCACAT----------------TCTGCCCC 4871

Qy     420 AGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
           | |||||||||| | |  ||||||||||||||| ||
Db    4872 ACAGAAAGTCCTTGGTTATCACCAGCAAGCTTGCAGG 4908


RESULT  22
AC023294
LOCUS       AC023294               98113 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-17L18, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC023294
VERSION     AC023294.2  GI:9123869
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 98113)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-17L18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 98113)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6960434.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3656
                Center clone name: 17_L_18
            --------------.
            * NOTE: This record contains 90 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *       1      996: contig of 996 bp in length
            *     997 1096: gap of      100 bp
            *    1097     2083: contig of 987 bp in length
            *    2084 2183: gap of      100 bp
            *    2184     3189: contig of 1006 bp in length
            *    3190 3289: gap of      100 bp
            *    3290     4298: contig of 1009 bp in length
            *    4299 4398: gap of      100 bp
            *    4399     5381: contig of 983 bp in length
            *    5382 5481: gap of      100 bp
            *    5482     6423: contig of 942 bp in length
            *    6424 6523: gap of      100 bp
            *    6524     7541: contig of 1018 bp in length
            *    7542 7641: gap of      100 bp
            *    7642     8604: contig of 963 bp in length
            *    8605 8704: gap of      100 bp
            *    8705     9669: contig of 965 bp in length
            *    9670 9769: gap of      100 bp
            *    9770    10781: contig of 1012 bp in length
            *   10782 10881: gap of      100 bp
            *   10882    11829: contig of 948 bp in length
            *   11830 11929: gap of      100 bp
            *   11930    12922: contig of 993 bp in length
            *   12923 13022: gap of      100 bp
            *   13023    14013: contig of 991 bp in length
            *   14014 14113: gap of      100 bp
            *   14114    15128: contig of 1015 bp in length
            *   15129 15228: gap of      100 bp
            *   15229    16225: contig of 997 bp in length
            *   16226 16325: gap of      100 bp
            *   16326    17325: contig of 1000 bp in length
            *   17326 17425: gap of      100 bp
            *   17426    18419: contig of 994 bp in length
            *   18420 18519: gap of      100 bp
            *   18520    19517: contig of 998 bp in length
            *   19518 19617: gap of      100 bp
            *   19618    20592: contig of 975 bp in length
            *   20593 20692: gap of      100 bp
            *   20693    21693: contig of 1001 bp in length
            *   21694 21793: gap of      100 bp
            *   21794    22784: contig of 991 bp in length
            *   22785 22884: gap of      100 bp
            *   22885    23897: contig of 1013 bp in length
            *   23898 23997: gap of      100 bp
            *   23998    24980: contig of 983 bp in length
            *   24981 25080: gap of      100 bp
            *   25081    26067: contig of 987 bp in length
            *   26068 26167: gap of      100 bp
            *   26168    27142: contig of 975 bp in length
            *   27143 27242: gap of      100 bp
            *   27243    28264: contig of 1022 bp in length
            *   28265 28364: gap of      100 bp
            *   28365    29384: contig of 1020 bp in length
            *   29385 29484: gap of      100 bp
            *   29485    30487: contig of 1003 bp in length
            *   30488 30587: gap of      100 bp
            *   30588    31595: contig of 1008 bp in length
            *   31596 31695: gap of      100 bp
            *   31696    32706: contig of 1011 bp in length
            *   32707 32806: gap of      100 bp
            *   32807    33747: contig of 941 bp in length
            *   33748 33847: gap of      100 bp
            *   33848    34860: contig of 1013 bp in length
            *   34861 34960: gap of      100 bp
            *   34961    35923: contig of 963 bp in length
            *   35924 36023: gap of      100 bp
            *   36024    36997: contig of 974 bp in length
            *   36998 37097: gap of      100 bp
            *   37098    38112: contig of 1015 bp in length
            *   38113 38212: gap of      100 bp
            *   38213    39249: contig of 1037 bp in length
            *   39250 39349: gap of      100 bp
            *   39350    40364: contig of 1015 bp in length
```

```
*    40365 40464: gap of      100 bp
*    40465    41450: contig of 986 bp in length
*    41451 41550: gap of      100 bp
*    41551    42542: contig of 992 bp in length
*    42543 42642: gap of      100 bp
*    42643    43654: contig of 1012 bp in length
*    43655 43754: gap of      100 bp
*    43755    44761: contig of 1007 bp in length
*    44762 44861: gap of      100 bp
*    44862    45833: contig of 972 bp in length
*    45834 45933: gap of      100 bp
*    45934    46886: contig of 953 bp in length
*    46887 46986: gap of      100 bp
*    46987    47948: contig of 962 bp in length
*    47949 48048: gap of      100 bp
*    48049    49034: contig of 986 bp in length
*    49035 49134: gap of      100 bp
*    49135    50128: contig of 994 bp in length
*    50129 50228: gap of      100 bp
*    50229    51198: contig of 970 bp in length
*    51199 51298: gap of      100 bp
*    51299    52314: contig of 1016 bp in length
*    52315 52414: gap of      100 bp
*    52415    53389: contig of 975 bp in length
*    53390 53489: gap of      100 bp
*    53490    54509: contig of 1020 bp in length
*    54510 54609: gap of      100 bp
*    54610    55578: contig of 969 bp in length
*    55579 55678: gap of      100 bp
*    55679    56672: contig of 994 bp in length
*    56673 56772: gap of      100 bp
*    56773    57768: contig of 996 bp in length
*    57769 57868: gap of      100 bp
*    57869    58834: contig of 966 bp in length
*    58835 58934: gap of      100 bp
*    58935    59929: contig of 995 bp in length
*    59930 60029: gap of      100 bp
*    60030    61038: contig of 1009 bp in length
*    61039 61138: gap of      100 bp
*    61139    62129: contig of 991 bp in length
*    62130 62229: gap of      100 bp
*    62230    63225: contig of 996 bp in length
*    63226 63325: gap of      100 bp
*    63326    64369: contig of 1044 bp in length
*    64370 64469: gap of      100 bp
*    64470    65464: contig of 995 bp in length
*    65465 65564: gap of      100 bp
*    65565    66575: contig of 1011 bp in length
*    66576 66675: gap of      100 bp
*    66676    67664: contig of 989 bp in length
*    67665 67764: gap of      100 bp
*    67765    68797: contig of 1033 bp in length
*    68798 68897: gap of      100 bp
*    68898    69897: contig of 1000 bp in length
*    69898 69997: gap of      100 bp
*    69998    71002: contig of 1005 bp in length
*    71003 71102: gap of      100 bp
*    71103    72121: contig of 1019 bp in length
*    72122 72221: gap of      100 bp
*    72222    73257: contig of 1036 bp in length
*    73258 73357: gap of      100 bp
*    73358    74351: contig of 994 bp in length
*    74352 74451: gap of      100 bp
*    74452    75462: contig of 1011 bp in length
*    75463 75562: gap of      100 bp

Query Match             23.9%;  Score 286.4;  DB 2;  Length 98113;
Best Local Similarity   81.6%;  Pred. No. 1.1e-69;
Matches  373;  Conservative    0;  Mismatches   67;  Indels   17;  Gaps    3;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           |||||||||||| ||||||||||  |||||||||||  |||||| |||||  |||||||  ||
Db   89944 GTGACAGGAGCCAAAGCAGCAGCATAGGTTGTCCCCATTTCCCTTCCCCTTTCCCTTGTC 90003

Qy      61 CGGTTGCCTTCCCGGGCCC-CTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAAC 119
           ||| | ||    |||||| | ||  | ||| || | | ||||  |||  |||||| |||
Db   90004 AGGTGGACTATATGGGCCCTCCTATGCCCCAGAGACTCAGTCCTACCACATCCCAGTAAC 90063

Qy     120 AAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAG 179
           |||||  |||||||||||| |||| |||||  | ||||||| | | | | |||||| ||||
Db   90064 AAGAAAGAGAGAACCCTGTGGAGAAGACCAACAAGGAGAAAC-GAAGATGCAGGATAAAG 90122

Qy     180 AAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAG 239
           ||||||||||  |||||||||| ||| ||| |||||||||||| |||||||||| ||||||||
Db   90123 AAGGTATTCTCTCTGAGAGAGTTGAGGAGCCAAAGCTAAAGTCCAAATACCTAAGCCTAG 90182

Qy     240 GACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACT 299
           ||||| |||| |||||||||| |||||||||||||||| | |||||||||| ||| |||||
Db   90183 GACAAGAGCCCGGAGGCTCCAACTTCCTCATGAAGAGATTACAGAAAGGGAAAAGGTACT 90242

Qy     300 TTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTG 359
           ||| ||||||||||||||||||||||||||||||||||||| |||||||||| |||||||
Db   90243 TTGTCTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAGTAAGCAGCTTCCAAGTG 90302

Qy     360 CAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCC 419
           ||| |||| | ||||||||||||||||||| |||||                ||||||||
Db   90303 CAGAACCAAAGAAGAACCTGGTGACTGGTGACCACAT---------------TCTGCCCC 90347

Qy     420 AGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
           | ||||||| || | |  ||||||||||||||| ||
Db   90348 ACAGAAAGTNCTTGGTTATCACCAGCAAGCTTGCAGG 90384


RESULT  23
AC091596
LOCUS       AC091596               69732 bp    DNA     linear   HTG 28-NOV-2001
DEFINITION  Homo sapiens chromosome 17 clone RP11-474K4 map 17, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC091596
VERSION     AC091596.2  GI:17047214
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 69732)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-474K4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 69732)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 22, 2001 this sequence version replaced gi:13959213.
            All repeats were identified using RepeatMasker:
```

```
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
    Center project name: L13230
    Center clone name: 474_K_4
--------------.
* NOTE: This record contains 87 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      703: contig of 703 bp in length
*      704 803: gap of      100 bp
*      804     1494: contig of 691 bp in length
*     1495 1594: gap of      100 bp
*     1595     2369: contig of 775 bp in length
*     2370 2469: gap of      100 bp
*     2470     3158: contig of 689 bp in length
*     3159 3258: gap of      100 bp
*     3259     3977: contig of 719 bp in length
*     3978 4077: gap of      100 bp
*     4078     4743: contig of 666 bp in length
*     4744 4843: gap of      100 bp
*     4844     5563: contig of 720 bp in length
*     5564 5663: gap of      100 bp
*     5664     6345: contig of 682 bp in length
*     6346 6445: gap of      100 bp
*     6446     7156: contig of 711 bp in length
*     7157 7256: gap of      100 bp
*     7257     7963: contig of 707 bp in length
*     7964 8063: gap of      100 bp
*     8064     8750: contig of 687 bp in length
*     8751 8850: gap of      100 bp
*     8851     9519: contig of 669 bp in length
*     9520 9619: gap of      100 bp
*     9620    10352: contig of 733 bp in length
*    10353 10452: gap of      100 bp
*    10453    11154: contig of 702 bp in length
*    11155 11254: gap of      100 bp
*    11255    11937: contig of 683 bp in length
*    11938 12037: gap of      100 bp
*    12038    12730: contig of 693 bp in length
*    12731 12830: gap of      100 bp
*    12831    13527: contig of 697 bp in length
*    13528 13627: gap of      100 bp
*    13628    14348: contig of 721 bp in length
*    14349 14448: gap of      100 bp
*    14449    15174: contig of 726 bp in length
*    15175 15274: gap of      100 bp
*    15275    16001: contig of 727 bp in length
*    16002 16101: gap of      100 bp
*    16102    16799: contig of 698 bp in length
*    16800 16899: gap of      100 bp
*    16900    17580: contig of 681 bp in length
*    17581 17680: gap of      100 bp
*    17681    18359: contig of 679 bp in length
*    18360 18459: gap of      100 bp
*    18460    19148: contig of 689 bp in length
*    19149 19248: gap of      100 bp
*    19249    19935: contig of 687 bp in length
*    19936 20035: gap of      100 bp
*    20036    20718: contig of 683 bp in length
*    20719 20818: gap of      100 bp
*    20819    21535: contig of 717 bp in length
*    21536 21635: gap of      100 bp
*    21636    22353: contig of 718 bp in length
*    22354 22453: gap of      100 bp
*    22454    23163: contig of 710 bp in length
*    23164 23263: gap of      100 bp
*    23264    23967: contig of 704 bp in length
*    23968 24067: gap of      100 bp
*    24068    24768: contig of 701 bp in length
*    24769 24868: gap of      100 bp
*    24869    25545: contig of 677 bp in length
*    25546 25645: gap of      100 bp
*    25646    26366: contig of 721 bp in length
*    26367 26466: gap of      100 bp
*    26467    27158: contig of 692 bp in length
*    27159 27258: gap of      100 bp
*    27259    27958: contig of 700 bp in length
*    27959 28058: gap of      100 bp
*    28059    28760: contig of 702 bp in length
*    28761 28860: gap of      100 bp
*    28861    29590: contig of 730 bp in length
*    29591 29690: gap of      100 bp
*    29691    30413: contig of 723 bp in length
*    30414 30513: gap of      100 bp
*    30514    31242: contig of 729 bp in length
*    31243 31342: gap of      100 bp
*    31343    32073: contig of 731 bp in length
*    32074 32173: gap of      100 bp
*    32174    32876: contig of 703 bp in length
*    32877 32976: gap of      100 bp
*    32977    33670: contig of 694 bp in length
*    33671 33770: gap of      100 bp
*    33771    34469: contig of 699 bp in length
*    34470 34569: gap of      100 bp
*    34570    35271: contig of 702 bp in length
*    35272 35371: gap of      100 bp
*    35372    36066: contig of 695 bp in length
*    36067 36166: gap of      100 bp
*    36167    36867: contig of 701 bp in length
*    36868 36967: gap of      100 bp
*    36968    37670: contig of 703 bp in length
*    37671 37770: gap of      100 bp
*    37771    38497: contig of 727 bp in length
*    38498 38597: gap of      100 bp
*    38598    39313: contig of 716 bp in length
*    39314 39413: gap of      100 bp
*    39414    40135: contig of 722 bp in length
*    40136 40235: gap of      100 bp
*    40236    40925: contig of 690 bp in length
*    40926 41025: gap of      100 bp
*    41026    41728: contig of 703 bp in length
*    41729 41828: gap of      100 bp
*    41829    42518: contig of 690 bp in length
*    42519 42618: gap of      100 bp
*    42619    43314: contig of 696 bp in length
*    43315 43414: gap of      100 bp
*    43415    44112: contig of 698 bp in length
*    44113 44212: gap of      100 bp
*    44213    44908: contig of 696 bp in length
*    44909 45008: gap of      100 bp
*    45009    45707: contig of 699 bp in length
*    45708 45807: gap of      100 bp
*    45808    46528: contig of 721 bp in length
*    46529 46628: gap of      100 bp
*    46629    47355: contig of 727 bp in length
*    47356 47455: gap of      100 bp
*    47456    48179: contig of 724 bp in length
*    48180 48279: gap of      100 bp
*    48280    48992: contig of 713 bp in length
*    48993 49092: gap of      100 bp
*    49093    49803: contig of 711 bp in length
*    49804 49903: gap of      100 bp
```

```
            *    49904    50605: contig of 702 bp in length
            *    50606 50705: gap of      100 bp
            *    50706    51397: contig of 692 bp in length
            *    51398 51497: gap of      100 bp
            *    51498    52196: contig of 699 bp in length
            *    52197 52296: gap of      100 bp
            *    52297    52999: contig of 703 bp in length
            *    53000 53099: gap of      100 bp
            *    53100    53787: contig of 688 bp in length
            *    53788 53887: gap of      100 bp
            *    53888    54545: contig of 658 bp in length
            *    54546 54645: gap of      100 bp
            *    54646    55332: contig of 687 bp in length

  Query Match             19.1%;  Score 229.4;  DB 2;  Length 69732;
  Best Local Similarity   83.6%;  Pred. No. 1.6e-53;
  Matches  290;  Conservative    0;  Mismatches   41;  Indels   16;  Gaps    2;

Qy     110 TCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACG 169
           | |||| ||||||||  |||||||||| |||| |||||  | ||||||| | | | | |
Db   24087 TTCCAGTAACAAGAAAGAGAGAACCCTGTGGAGAAGACCAACAAGGAGAAAC-GAAGATG 24145

Qy     170 CAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATAC 229
           ||||| ||||||||||||||  |||||||| ||| ||| |||||||||| ||||||||
Db   24146 CAGGATAAAGAAGGTATTCTCTCTGAGAGAGTTGAGGAGCCAAAGCTAAAGTCCAAATAC 24205

Qy     230 CCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGG 289
           | |||||||||||| |||| ||||||||| |||||||||||||||| | |||||||||
Db   24206 CTAAGCCTAGGACAAGAGCCCGGAGGCTCCAACTTCCTCATGAAGAGATTACAGAAAGGG 24265

Qy     290 CAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAG 349
            ||| |||||||| |||||||||||||||||||||||||||||||||||| |||||||
Db   24266 AAAAGGTACTTTGTCTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAGTAAGCAG 24325

Qy     350 CTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAG 409
           || |||||||||| |||| | ||||||||||||||||||| |||||  .
Db   24326 CTTCCAAGTGCAGAACCAAAGAAGAACCTGGTGACTGGTGACCACAT------------ 24372

Qy     410 GATCTGCCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
             ||||||||| |||||||||| | |  ||||||||||||||| ||
Db   24373 --TCTGCCCCACAGAAAGTCCTTGGTTATCACCAGCAAGCTTGCAGG 24417


RESULT  24
BTAENSULF
LOCUS       BTAENSULF               231 bp    mRNA    linear   MAM 09-MAY-1998
DEFINITION  B.taurus mRNA for alpha endosulfine.
ACCESSION   X95958
VERSION     X95958.1  GI:1502303
KEYWORDS    alpha endosulfine.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 231)
  AUTHORS   Peyrollier,K., Heron,L., Virsolvy-Vergine,A., Le Cam,A. and
            Bataille,D.
  TITLE     Alpha endosulfine is a novel molecule, structurally related to a
            family of phosphoproteins
  JOURNAL   Biochem. Biophys. Res. Commun. 223 (3), 583-586 (1996)
  MEDLINE   96268722
REFERENCE   2  (bases 1 to 231)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-1996) D. Bataille, INSERM U 376, CHU
            Arnaud-de-Villeneuve, 371 Rue du Doyen Gaston Giraud, F- 34295
            Montpellier Cedex 5, FRANCE
COMMENT     Related sequence:
            Virsolvy-Vergine, Diabetologia 39, 135-141, 1996.
FEATURES             Location/Qualifiers
     source          1. .231
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /dev_stage="adult"
     CDS             <1. .>231
                     /codon_start=1
                     /product="alpha endosulfine"
                     /protein_id="CAA65196.1"
                     /db_xref="GI:1502304"
                     /db_xref="SPTREMBL:Q95105"
                     /translation="EGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSG
                     DYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQ"
BASE COUNT       81 a     58 c     57 g     35 t
ORIGIN


  Query Match             17.8%;  Score 213.4;  DB 4;  Length 231;
  Best Local Similarity   95.2%;  Pred. No. 1.8e-49;
  Matches  220;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy     179 GAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTA 238
           ||||||||||||||||||||||||||| |||||||||||||| |||||||| |||||||||
Db       1 GAAGGTATTCTGCCTGAGAGAGCTGAGGAGGCAAAGCTAAAAGCCAAATATCCAAGCCTA 60

Qy     239 GGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTAC 298
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      61 GGACAAAAGCCTGGAGGTTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTAC 120

Qy     299 TTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGT 358
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||| || |||
Db     121 TTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAACAAGCAGCTACCGAGT 180

Qy     359 GCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAG 409
           |||||||||||||||||||| |||||||| ||||| ||||| ||||||
Db     181 GCAGGACCAGACAAGAACCTGGTCACTGGTGACCACATTCCCACTCCACAG 231


RESULT  25
AC092479/c
LOCUS       AC092479               142023 bp    DNA     linear   HTG 05-FEB-2002
DEFINITION  Mus musculus clone rp23-231l15 strain C57BL/6J, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC092479
VERSION     AC092479.12  GI:18497125
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 142023)
  AUTHORS   Jia,H., Zhang,P., Lin,S., Wu,H. and Roe,B.A.
  TITLE     Mus musculus BAC Clone rp23-231l15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 142023)
  AUTHORS   Jia,H., Zhang,P., Lin,S., Wu,H. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2001) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Feb 5, 2002 this sequence version replaced gi:18390246.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    32949: contig of 32949 bp in length
            *    32950    33049: gap of unknown length
```

```
           *    33050    78918: contig of 45869 bp in length
           *    78919    79018: gap of unknown length
           *    79019   142023: contig of 63005 bp in length.
FEATURES             Location/Qualifiers
     source          1. .142023
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="rp23-231l15"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    37523 a  34071 c  31510 g  38713 t    206 others
ORIGIN


  Query Match             15.4%;  Score 185;  DB 2;  Length 142023;
  Best Local Similarity   65.1%;  Pred. No. 7.7e-41;
  Matches  420;  Conservative    0;  Mismatches  175;  Indels   50;  Gaps    8;

Qy     456 GTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGC 515
           ||||||||||||| ||||||| || ||||||| |||||||| ||||||
Db  135003 GTGGCCAAGTTGAGTGATGCTTCCTGGGGCTCCACCAGATCGTGAGACGC---------- 134952

Qy     516 TTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCCTTCAGCCAGGGGTCAG 575
           | ||| | | |  ||      | |   |||     |||||   |  ||||||||||||
Db  134953 -TGCCCACTCGGGTCCTGTGCTGGCTCCTGCCCCCTCCCTGCTTTTGTAGCCAGGGGTCAG 134895

Qy     576 GAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACA 635
           |||||||| ||||||  |||| |||||||||| ||  | | ||   |||||||||| ||||
Db  134894 GAGGTGGCCCGGGTGCAGGCTAGAGAGGCAG-AGAACCTGCCCACTGGTGTCCCAACACA 134836

Qy     636 TGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTT-------GTTTTACCTTT 688
           ||||||||||||||||||||||||||||||||||||||||| ||||||         |||  | |||
Db  134835 TGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACTTTTTTTTGTTTTGTTTTACCTTTT 134776

Qy     689 TTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGTGGGGCTGAA 748
           |||||||||||| ||||| |||||  ||| ||| || |          |||||||| ||||
Db  134775 TTTCCAAATAACTGTTGGAAGAAAACTCATTGAGATCC------CAAGGGTGGGGTTGAA 134722

Qy     749 AGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACATTC 808
           ||| ||| |||    | |||  || |  ||| | |||| ||     || ||||||
Db  134721 AGGATGGAGTG---AAGATGTGGGGGAGTGAGTGAGGGTTTATGACTGATAAAAAAGAAG 134665

Qy     809 CTGACTATCCTTCTTAACCACGTGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAGTGGA 868
           |  ||| | ||||     || |               ||  ||  | |   ||     |
Db  134664 TTTCCTACCGTTCTCCTGAACCT------------CTGGGTTGCTGTGGTTGACTGTAA 134618

Qy     869 GTAGCCTAATGAAAAGGGGTTCTAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAGTGTG 928
           |  |    |||  |||||||||  ||| ||||||| ||| | | ||  | ||| | | ||
Db  134617 GGTGAGGGATGGTAAGGGGTTCAGGTTCAGCTCTGAAGAGAGAAGCACTCTTTGAATATG 134558

Qy     929 GTTGGAGACCTGGTGTCAGATAAAAGAAACTCCATCCGCACAGACAGATGCAAACAGCTC 988
           | | | || || |||||||| ||||| || |   | |    |  |||| |    |
Db  134557 GCTAGTGATCTAGTGTCAGAGAAAAGGAATTGTGTTCTTTAAAGCAGACGAGGTCT---- 134500

Qy     989 CTCTAGTTCTGCAGAGCTAGTTGAGAACTCAACATTAAT-----CATTTTAAAAAGTACT 1043
           |||| |||||||||| | ||||||| | | |  || ||      | ||| |||| | ||
Db  134501 CTCTGGTTCTGCAGAACAAGTTGAGGATTTAGGATAAACTGTTCCTTTTAAAAAAATGCT 134442

Qy    1044 GTCCTTGAAATAGATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGG 1088
           |||||||||||| ||||||||| | |||||||||||||| || ||
Db  134441 GTCCTTGAAATAGGTTTGCTGTGTGGAGAAGGGCAGTGAATGGGG 134397


RESULT  26
MMA005983
LOCUS       MMA005983               339 bp    mRNA    linear   ROD 07-JAN-1999
DEFINITION  Mus musculus mRNA for cAMP-regulated phosphoprotein (ARPP-19).
ACCESSION   AJ005983
VERSION     AJ005983.1  GI:4138189
KEYWORDS    ARPP-19 gene; cAMP-regulated phosphoprotein.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Bataille D., U 376 'Endocrinologie des
            Peptides et regulation genique', INSERM, CHU Arnaud-de-Villeneuve,
            34295, FRANCE
REFERENCE   2  (bases 1 to 339)
  AUTHORS   Girault,J.A., Horiuchi,A., Gustafson,E., Rosen,N. and Greengard,P.
  TITLE     Differential expression of ARPP-16 and ARPP-19, two highly related
            phosphoproteins, one of which is specifically associated with
            dopamine-innervated brain regions
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /tissue_type="brain"
     gene            1. .339
                     /gene="ARPP-19"
     CDS             1. .339
                     /gene="ARPP-19"
                     /codon_start=1
                     /product="cAMP-regulated phosphoprotein"
                     /protein_id="CAA06797.1"
                     /db_xref="GI:4138190"
                     /translation="MSAEVPEAASAEEQKEMEDKVTSPEKAEEAKLKARYPHLGQKPG
                     GSDFLRKRLQKGQKYFDSGDYNMAKAKMKNKQLPAAAPDKTEVTGDHIPTPQDLPQRK
                     PSLVASKLAG"
BASE COUNT      111 a     73 c     91 g     64 t
ORIGIN


  Query Match             14.6%;  Score 175.4;  DB 10;  Length 339;
  Best Local Similarity   72.9%;  Pred. No. 1.2e-38;
  Matches  240;  Conservative    0;  Mismatches   86;  Indels    3;  Gaps    1;

Qy     128 GAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATT 187
           ||   ||| | ||  | | | |  |||||| |  |||| | | | || |||| |  || |
Db      10 GAAGTCCCCGAGGCAGCGTCTGCGGAGGAGCAAAAGGAAATGGAAGATAAAGTAACTAGT 69

Qy     188 CTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAG 247
              || |||| |||||||||| |||||| ||||||| |  || ||    | | |||||||||
Db      70 ---CCAGAGAAAGCTGAAGAAGCAAAGTTAAAGGCAAGGTATCCTCACTTGGGACAAAAG 126

Qy     248 CCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCA 307
           |||||||| || || ||   | | ||| ||| | ||||||||||||||||| ||||| ||
Db     127 CCTGGAGGGTCTGATTTTTTAAGGAAAAGATTGCAGAAAGGGCAAAAGTATTTTGATTCT 186

Qy     308 GGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCA 367
           || |||||||||||||| ||||| |||||||||| |||||||| ||   |||||  ||
Db     187 GGGGACTACAACATGGCAAAAGCAAAGATGAAGAACAAGCAGCTTCCTGCTGCAGCCCCG 246

Qy     368 GACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAG 427
           || ||||    ||| |||||||| ||||| ||||| |||||||| || || ||| | ||
Db     247 GATAAGACAGAGGTCACTGGTGACCACATTCCCACTCCACAGGACCTTCCTCAGCGGAAA 306

Qy     428 TCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
           | || || ||  | ||||| || || ||
Db     307 CCATCCCTTGTTGCTAGCAAACTGGCTGG 335


RESULT  27
AC103421
LOCUS       AC103421               105126 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-148B7, *** SEQUENCING IN PROGRESS
            ***, 47 unordered pieces.
ACCESSION   AC103421
VERSION     AC103421.2  GI:17973226
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 105126)

AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 105126)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (25-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Dec 20, 2001 this sequence version replaced gi:17064207.

---------- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

---------- Project Information

Center project name: GHMB
Center clone name: CH230-148B7

---------- Summary Statistics

Assembly program: Phrap; version 0.990329First call to findPhrapList
Consensus quality: 87267 bases at least Q40
Consensus quality: 93895 bases at least Q30
Consensus quality: 100375 bases at least Q20
Estimated insert size: 85832; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1x in Q20 bases; sum-of-contigs estimation

---

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 47 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	4645: contig of 4645 bp in length
*	4646	4745: gap of unknown length
*	4746	8474: contig of 3729 bp in length
*	8475	8574: gap of unknown length
*	8575	12295: contig of 3721 bp in length
*	12296	12395: gap of unknown length
*	12396	17604: contig of 5209 bp in length
*	17605	17704: gap of unknown length
*	17705	21282: contig of 3578 bp in length
*	21283	21382: gap of unknown length
*	21383	24900: contig of 3518 bp in length
*	24901	25000: gap of unknown length
*	25001	27522: contig of 2522 bp in length
*	27523	27622: gap of unknown length
*	27623	30513: contig of 2891 bp in length
*	30514	30613: gap of unknown length
*	30614	34148: contig of 3535 bp in length
*	34149	34248: gap of unknown length
*	34249	37150: contig of 2902 bp in length
*	37151	37250: gap of unknown length
*	37251	40253: contig of 3003 bp in length
*	40254	40353: gap of unknown length
*	40354	42860: contig of 2507 bp in length
*	42861	42960: gap of unknown length
*	42961	45644: contig of 2684 bp in length
*	45645	45744: gap of unknown length
*	45745	47445: contig of 1701 bp in length
*	47446	47545: gap of unknown length
*	47546	49806: contig of 2261 bp in length
*	49807	49906: gap of unknown length
*	49907	53212: contig of 3306 bp in length
*	53213	53312: gap of unknown length
*	53313	56240: contig of 2928 bp in length
*	56241	56340: gap of unknown length
*	56341	59832: contig of 3492 bp in length
*	59833	59932: gap of unknown length
*	59933	62010: contig of 2078 bp in length
*	62011	62110: gap of unknown length
*	62111	63482: contig of 1372 bp in length
*	63483	63582: gap of unknown length
*	63583	65587: contig of 2005 bp in length
*	65588	65687: gap of unknown length
*	65688	67278: contig of 1591 bp in length
*	67279	67378: gap of unknown length
*	67379	69837: contig of 2459 bp in length
*	69838	69937: gap of unknown length
*	69938	71641: contig of 1704 bp in length
*	71642	71741: gap of unknown length
*	71742	72772: contig of 1031 bp in length
*	72773	72872: gap of unknown length
*	72873	74038: contig of 1166 bp in length
*	74039	74138: gap of unknown length
*	74139	75347: contig of 1209 bp in length
*	75348	75447: gap of unknown length
*	75448	76582: contig of 1135 bp in length
*	76583	76682: gap of unknown length
*	76683	78565: contig of 1883 bp in length
*	78566	78665: gap of unknown length
*	78666	80058: contig of 1393 bp in length
*	80059	80158: gap of unknown length
*	80159	81800: contig of 1642 bp in length
*	81801	81900: gap of unknown length
*	81901	83526: contig of 1626 bp in length
*	83527	83626: gap of unknown length
*	83627	85206: contig of 1580 bp in length
*	85207	85306: gap of unknown length
*	85307	86373: contig of 1067 bp in length

```
            *    86374    86473: gap of unknown length
            *    86474    88280: contig of 1807 bp in length
            *    88281    88380: gap of unknown length
            *    88381    90567: contig of 2187 bp in length
            *    90568    90667: gap of unknown length
            *    90668    91832: contig of 1165 bp in length
            *    91833    91932: gap of unknown length
            *    91933    92951: contig of 1019 bp in length
            *    92952    93051: gap of unknown length
            *    93052    94177: contig of 1126 bp in length
            *    94178    94277: gap of unknown length
            *    94278    95584: contig of 1307 bp in length
            *    95585    95684: gap of unknown length
            *    95685    96966: contig of 1282 bp in length
            *    96967    97066: gap of unknown length
            *    97067    98319: contig of 1253 bp in length
            *    98320    98419: gap of unknown length
            *    98420    99500: contig of 1081 bp in length
            *    99501    99600: gap of unknown length
            *    99601   100809: contig of 1209 bp in length
            *   100810   100909: gap of unknown length
            *   100910   102534: contig of 1625 bp in length
            *   102535   102634: gap of unknown length
            *   102635   103650: contig of 1016 bp in length
            *   103651   103750: gap of unknown length
            *   103751   105126: contig of 1376 bp in length.
FEATURES             Location/Qualifiers
     source          1. .105126
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-148B7"
BASE COUNT    31209 a   19023 c   19380 g   30766 t    4748 others
ORIGIN
```

Query Match 14.3%; Score 172; DB 2; Length 105126;
Best Local Similarity 75.4%; Pred. No. 3.6e-37;
Matches 214; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```
Qy     173 GAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCA 232
           ||| |||       |||    | |||||  || || |  ||     | ||||| ||| |||
Db   97782 GAGGGAGAGAAGGATCTAGGAGTGAGAGAGGATGAAGTTAAAAGGAGGGCCATATATCCA 97841

Qy     233 AGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAA 292
           |||| ||||||||||||||||||||  | ||||||| ||||||||  || |||||||
Db   97842 CATCTAGAACAAAAGCCTGGAGGCTCCATCATCCTCATGGAGAGACTCTTGATAGGGCAA 97901

Qy     293 AAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTG 352
           || |||||||| |||||||||| | || |  ||||||||||||||||  |||||| ||
Db   97902 AATTACTTTGAATCAGGAGACTATACCAAGTTCAAAGCCAAGATGAAGATCAAGCAGATG 97961

Qy     353 CCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGAT 412
           |  ||||||||  |||||| |||| || ||||  |  | | |||||| | | ||||||||
Db   97962 CTGAGTGCAGGGGCAGACATGAACTTGATGACCAGCAACCTCATCCCTATCACACAGGAT 98021

Qy     413 CTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
           |||||||||||| ||| ||| ||  ||| ||||||||||| |||
Db   98022 CTGCCCCAGAGAGAGTTCTCACTTATCATCAGCAAGCTTGTGGG 98065
```

```
RESULT  28
RNA005982
LOCUS       RNA005982               339 bp    mRNA    linear   ROD 07-JAN-1999
DEFINITION  Rattus norvegicus mRNA for cAMP-regulated phosphoprotein (ARPP-19).
ACCESSION   AJ005982
VERSION     AJ005982.1  GI:4138423
KEYWORDS    ARPP-19 gene; cAMP-regulated phosphoprotein.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Bataille D., U 376 'Endocrinologie des
            Peptides et regulation genique', INSERM, CHU Arnaud-de-Villeneuve,
            34295, FRANCE
REFERENCE   2  (bases 1 to 339)
  AUTHORS   Girault,J.A., Horiuchi,A., Gustafson,E., Rosen,N. and Greengard,P.
  TITLE     Differential expression of ARPP-16 and ARPP-19, two highly related
            cAMP-regulated phosphoproteins, one of which is specifically
            associated with dopamine-innervated brain regions
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /tissue_type="brain"
     gene            1. .339
                     /gene="ARPP-19"
     CDS             1. .339
                     /gene="ARPP-19"
                     /codon_start=1
                     /product="cAMP-regulated phosphoprotein"
                     /protein_id="CAA06796.1"
                     /db_xref="GI:4138424"
                     /translation="MSAEVPEAASAEEQKEMEDKVTSPEKAEEAKLKARYPHLGQKPG
                     GSDFLRKRLQKGQKYFDSGDYNMAKAKMKNKQLPAAAPDKTEVTGDHIPTPQDLPQRK
                     PSLVASKLAG"
BASE COUNT      112 a     72 c     89 g     66 t
ORIGIN
```

Query Match 14.2%; Score 170.2; DB 10; Length 339;
Best Local Similarity 70.8%; Pred. No. 3.7e-37;
Matches 226; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

```
Qy     138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
           | ||||   |    | |||| ||||  |      |||  |  |||  | |   || ||||
Db      17 CCGAGGCAGCGTCGGCGGAGGAGCAAAAAGAAATGGAAGATAAAGTAACTAGTCCAGAGA 76

Qy     198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
           |||||||| |||||| |||| || |  || ||     | ||||||||||||| || |
Db      77 AAGCTGAAGAAGCAAAGTTAAAAGCAAGGTATCCTCATTTGGGACAAAAGCCTGGTGGTT 136

Qy     258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
           | |||||  | | ||| ||| | ||||||||||||||||| |||||  || || ||||
Db     137 CTGACTTTTTAAGGAAAAGATTGCAGAAAGGGCAAAAGTATTTTGATTCTGGGGATTACA 196

Qy     318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
           |||||| |||| |||||||||| |||||||  ||   |||||  || || ||||
Db     197 ACATGGCAAAAGCAAAGATGAAGAACAAGCAGCTTCCTGCTGCAGCCCCGGATAAGACAG 256

Qy     378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
           ||| |||||||| |||| |||| |||||||| || || ||| | ||  | || || |
Db     257 AGGTCACTGGTGACCACATTCCCACTCCACAGGACCTTCCTCAGCGGAAACCATCCCTTG 316

Qy     438 TCACCAGCAAGCTTGCGGG 456
           |  | ||||| || || ||
Db     317 TTGCTAGCAAACTGGCTGG 335
```

```
RESULT  29
HOS223091
LOCUS       HOS223091               339 bp    mRNA    linear   PRI 22-SEP-1998
DEFINITION  Homo sapiens mRNA for ARPP-19 protein.
ACCESSION   AJ223091
VERSION     AJ223091.1  GI:3647145
KEYWORDS    ARPP-19 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Heron,L.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (18-DEC-1997) Heron L., Endocrinologie des peptides,
            INSERM U 376, CHU Arnaud-de-Villeneuve, 34295 Montpellier, FRANCE
REFERENCE   2  (bases 1 to 339)
  AUTHORS   Heron,L., Virsolvy,A., Peyrollier,K., Gribble,F.M., Le Cam,A.,
            Ashcroft,F.M. and Bataille,D.
  TITLE     Human alpha-endosulfine, a possible regulator of
            sulfonylurea-sensitive KATP channel: molecular cloning, expression
            and biological properties
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (14), 8387-8391 (1998)
  MEDLINE   98318667
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="brain"
     gene            1. .339
                     /gene="ARPP-19"
     CDS             1. .339
                     /gene="ARPP-19"
                     /codon_start=1
                     /protein_id="CAA11115.1"
                     /db_xref="GI:3647146"
                     /translation="MSAEVPEAASAEEQKEMEDKVTSPEKAEEAKLKARYPHLGQKPG
                     GSDFLRKRLQKGQKYFDSGDYNMAKAKMKNKQLPTAAPDKTEVTGDHIPTPQDLPQRK
                     PSLVASKLAG"
BASE COUNT      112 a     76 c     87 g     64 t
ORIGIN
```

```
  Query Match             14.1%;  Score 168.6;  DB 9;  Length 339;
  Best Local Similarity   70.5%;  Pred. No. 1e-36;
  Matches  225;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy     138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
           | ||||   ||  || |||| ||||| |      ||| | ||| | | |  || ||||
Db      17 CCGAGGCAGCCTCCGCGGAGGAGCAGAAGGAAATGGAAGATAAAGTGACTAGTCCAGAGA 76

Qy     198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
           ||| ||||| |||||  |||| || | ||| ||    || ||||||||||||||||| |
Db      77 AAGCAGAAGAAGCAAAATTAAAAGCAAGATATCCTCATCTGGGACAAAAGCCTGGAGGTT 136

Qy     258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
           |||| ||| | | ||| | | ||||||||||||||| || ||||| || || || ||||
Db     137 CCGATTTCTTAAGGAAACGGTTGCAGAAAGGGCAAAAATATTTTGATTCTGGGGATTACA 196

Qy     318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
           ||||||| ||||| || |||||||| ||||| || || | |||||  || || ||||
Db     197 ACATGGCTAAAGCAAAAATGAAGAACAAGCAACTTCCTACTGCAGCTCCGGATAAGACGG 256

Qy     378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
           ||| |||||||| ||||| ||||| || || || || || ||  | ||| | || || |
Db     257 AGGTCACTGGTGACCACATTCCCACTCCGCAAGACCTTCCTCAACGGAAGCCGTCCCTTG 316

Qy     438 TCACCAGCAAGCTTGCGGG 456
           |  | ||||| || || ||
Db     317 TTGCTAGCAAACTGGCTGG 335
```

```
RESULT  30
BOVARPP19
LOCUS       BOVARPP19               1188 bp    mRNA    linear   MAM 26-APR-1993
DEFINITION  Bovine cAMP-regulated phosphoprotein (ARPP-19) mRNA, complete cds.
ACCESSION   M33618 J05450
VERSION     M33618.1  GI:162689
KEYWORDS    ARPP-19; cAMP-regulated phosphoprotein.
SOURCE      Bovine young adult caudate nuclei, cDNA to mRNA, clone 1.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1188)
  AUTHORS   Horiuchi,A., Williams,K.R., Kurihara,T., Nairn,A.C. and
            Greengard,P.
  TITLE     Purification and cDNA cloning of ARPP-16, a cAMP-regulated
            phosphoprotein enriched in basal ganglia, and of a related
            phosphoprotein, ARPP-19
  JOURNAL   J. Biol. Chem. 265, 9476-9484 (1990)
  MEDLINE   90264450
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by A.Horiuchi, 06-APR-1990.
FEATURES             Location/Qualifiers
     source          1. .1188
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
     mRNA            1. .1188
                     /product="ARPP-19 mRNA"
     CDS             97. .435
                     /note="cAMP-regulated phosphoprotein"
                     /codon_start=1
                     /protein_id="AAA30386.1"
                     /db_xref="GI:162690"
                     /translation="MSAEVPEAASAEEQKEMEDKVTSPEKAEEAKLKARYPHLGQKPG
                     GSDFLRKRLQKGQKYFDSGDYNMAKAKMKNKQLPTATPDKTEVTGDHIPTPQDLPQRK
                     PSLVASKLAG"
BASE COUNT      349 a    211 c    261 g    367 t
ORIGIN
```

```
  Query Match             14.1%;  Score 168.6;  DB 4;  Length 1188;
  Best Local Similarity   70.5%;  Pred. No. 1.3e-36;
  Matches  225;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy     138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
           | ||||   ||  || |||| ||||| |      ||| | ||| | | |  || ||||
Db     113 CCGAGGCAGCCTCCGCGGAGGAGCAGAAGGAAATGGAAGATAAAGTGACTAGTCCAGAGA 172

Qy     198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
           ||||||||| |||||  |||| || |  || ||    || ||||||||||||||||| |
Db     173 AAGCTGAAGAAGCAAAATTAAAAGCAAGGTATCCTCATCTGGGACAAAAGCCTGGAGGTT 232

Qy     258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
           |||| ||  | | |||  || | ||||||||||||||| || ||||| || || || ||||
Db     233 CCGATTTTTTAAGGAAACGATTGCAGAAAGGGCAAAAATATTTTGATTCTGGGGATTACA 292

Qy     318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
           ||||||| ||||| || |||||||| ||||| || || | ||||   || || ||||
Db     293 ACATGGCTAAAGCAAAAATGAAGAACAAGCAACTTCCTACTGCAACCCCGGATAAGACAG 352

Qy     378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
           ||| |||||||| ||||| ||||| |||||||| || || ||  | ||  | || || |
Db     353 AGGTCACTGGTGACCACATTCCCACTCCACAGGACCTTCCTCAACGGAAACCATCTCTTG 412

Qy     438 TCACCAGCAAGCTTGCGGG 456
           |  | ||||| || || ||
Db     413 TTGCTAGCAAACTGGCTGG 431
```

```
RESULT  31
BC003418
LOCUS       BC003418                1228 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, cyclic AMP phosphoprotein, 19 kD, clone MGC:5468
            IMAGE:3451558, mRNA, complete cds.
ACCESSION   BC003418
VERSION     BC003418.1  GI:13097335
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
```

```
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalon@bcm.tmc.edu.
            Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
            A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
            Muzny,D.M., Gibbs,R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 3 Row: p Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="Homo sapiens"
                     /db_xref="LocusID:10776"
                     /db_xref="taxon:9606"
                     /clone="MGC:5468 IMAGE:3451558"
                     /tissue_type="Placenta, choriocarcinoma"
                     /clone_lib="NIH_MGC_10"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             125. .463
                     /codon_start=1
                     /product="cyclic AMP phosphoprotein, 19 kD"
                     /protein_id="AAH03418.1"
                     /db_xref="GI:13097336"
                     /translation="MSAEVPEAASAEEQKEMEDKVTSPEKAEEAKLKARYPHLGQKPG
                     GSDFLRKRLQKGQKYFDSGDYNMAKAKMKNKQLPTAAPDKTEVTGDHIPTPQDLPQRK
                     PSLVASKLAG"
BASE COUNT      354 a    229 c    272 g    373 t
ORIGIN


  Query Match             14.1%;  Score 168.6;  DB 9;  Length 1228;
  Best Local Similarity   70.5%;  Pred. No. 1.4e-36;
  Matches  225;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy    138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
          | ||||   ||  || |||| ||||| |      |||  |  |||  | |   || ||||
Db    141 CCGAGGCAGCCTCCGCGGAGGAGCAGAAGGAAATGGAAGATAAAGTGACTAGTCCAGAGA 200

Qy    198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
           ||| ||||| |||||  |||| || | ||| ||    || ||||||||||||||||| |
Db    201 AAGCAGAAGAAGCAAAATTAAAAGCAAGATATCCTCATCTGGGACAAAAGCCTGGAGGTT 260

Qy    258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
          | || ||| | | |||  |  | ||||||||||||||| || ||||| || || || ||||
Db    261 CAGATTTCTTAAGGAAACGGTTGCAGAAAGGGCAAAAATATTTTGATTCTGGGGATTACA 320

Qy    318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
          ||||||| ||||| || |||||||| ||||| || || | |||||  || || ||||
Db    321 ACATGGCTAAAGCAAAAATGAAGAACAAGCAACTTCCTACTGCAGCTCCGGATAAGACGG 380

Qy    378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
           ||| |||||||| ||||| ||||| || || || || || ||  | ||| | || || |
Db    381 AGGTCACTGGTGACCACATTCCCACTCCGCAAGACCTTCCTCAACGGAAGCCGTCCCTTG 440

Qy    438 TCACCAGCAAGCTTGCGGG 456
          |  | ||||||| || ||
Db    441 TTGCTAGCAAGCTGGCTGG 459
```

```
RESULT  32
AF084555
LOCUS       AF084555                5171 bp    mRNA    linear   PRI 16-AUG-2001
DEFINITION  Homo sapiens okadaic acid-inducible and cAMP-regulated
            phosphoprotein 19 (ARPP-19) mRNA, complete cds.
ACCESSION   AF084555
VERSION     AF084555.1  GI:5813858
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5171)
  AUTHORS   Chin,L.S., Singh,S.K., Wang,Q. and Murray,S.F.
  TITLE     Identification of okadaic-acid-induced genes by mRNA differential
            display in glioma cells
  JOURNAL   J. Biomed. Sci. 7 (2), 152-159 (2000)
  MEDLINE   20219739
   PUBMED   10754390
REFERENCE   2  (bases 1 to 5171)
  AUTHORS   Singh,S.K., Murray,S.F. and Chin,L.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-1998) Neurosurgery, University of Maryland at
            Baltimore, 685 W. Baltimore Street, Baltimore, MD 21201, USA
FEATURES             Location/Qualifiers
     source          1. .5171
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="OA4-5"
                     /tissue_type="brain"
                     /dev_stage="fetus"
     gene            1. .5171
                     /gene="ARPP-19"
     CDS             125. .463
                     /gene="ARPP-19"
                     /codon_start=1
                     /product="okadaic acid-inducible and cAMP-regulated
                     phosphoprotein 19"
                     /protein_id="AAD52044.1"
                     /db_xref="GI:5813859"
                     /translation="MSAEVPEAASAEEQKEMEDKVTSPEKAEEAKLKARYPHLGQKPG
                     GSDFLRKRLQKGQKYFDSGDYNMAKAKMKNKQLPTAAPDKTEVTGDHIPTPQDLPQRK
                     PSLVASKLAG"
BASE COUNT     1494 a    974 c    956 g    1746 t      1 others
ORIGIN


  Query Match             14.1%;  Score 168.6;  DB 9;  Length 5171;
  Best Local Similarity   70.5%;  Pred. No. 1.8e-36;
  Matches  225;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy    138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
          | ||||   ||  || |||| ||||| |      |||  |  |||  | |   || ||||
Db    141 CCGAGGCAGCCTCCGCGGAGGAGCAGAAGGAAATGGAAGATAAAGTGACTAGTCCAGAGA 200

Qy    198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
           ||| ||||| |||||  |||| || | ||| ||    || ||||||||||||||||| |
Db    201 AAGCAGAAGAAGCAAAATTAAAAGCAAGATATCCTCATCTGGGACAAAAGCCTGGAGGTT 260

Qy    258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
          | || ||| | | |||  |  | ||||||||||||||| || ||||| || || || ||||
Db    261 CAGATTTCTTAAGGAAACGGTTGCAGAAAGGGCAAAAATATTTTGATTCTGGGGATTACA 320

Qy    318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
          ||||||| ||||| || |||||||| ||||| || || | |||||  || || ||||
Db    321 ACATGGCTAAAGCAAAAATGAAGAACAAGCAACTTCCTACTGCAGCTCCGGATAAGACGG 380

Qy    378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
           ||| |||||||| ||||| ||||| || || || || || ||  | ||| | || || |
Db    381 AGGTCACTGGTGACCACATTCCCACTCCGCAAGACCTTCCTCAACGGAAGCCGTCCCTTG 440

Qy    438 TCACCAGCAAGCTTGCGGG 456
          |  | ||||||| || ||
```

```
Db     441 TTGCTAGCAAGCTGGCTGG 459


RESULT  33
AC091874
LOCUS       AC091874               71941 bp    DNA     linear   HTG 09-JUN-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2593A12, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC091874
VERSION     AC091874.1  GI:14333810
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 71941)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71941)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 825705
            Center clone name: CITB-E1_2593A12
            -------
            Summary Statistics
            Consensus quality: 62238 bases at least Q40
            Consensus quality: 67347 bases at least Q30
            Consensus quality: 68096 bases at least Q20
            Estimated insert size: 219000; agarose-fp estimation
            Estimated insert size: 71141; sum-of-contigs estimation
            Quality coverage:  4.14 in Q20 bases; agarose-fp estimation
            Quality coverage:  12.76 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1461: contig of 1461 bp in length
            *     1462     1561: gap of unknown length
            *     1562     2633: contig of 1072 bp in length
            *     2634     2733: gap of unknown length
            *     2734     4507: contig of 1774 bp in length
            *     4508     4607: gap of unknown length
            *     4608     7249: contig of 2642 bp in length
            *     7250     7349: gap of unknown length
            *     7350    12208: contig of 4859 bp in length
            *    12209    12308: gap of unknown length
            *    12309    20679: contig of 8371 bp in length
            *    20680    20779: gap of unknown length
            *    20780    29687: contig of 8908 bp in length
            *    29688    29787: gap of unknown length
            *    29788    39781: contig of 9994 bp in length
            *    39782    39881: gap of unknown length
            *    39882    71941: contig of 32060 bp in length.
FEATURES             Location/Qualifiers
     source          1. .71941
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2593A12"
                     /clone_lib="CalTech human BAC library D"

BASE COUNT    20351 a  15896 c  16000 g  18892 t    802 others
ORIGIN


  Query Match           13.9%;  Score 167.2;  DB 2;  Length 71941;
  Best Local Similarity  67.5%;  Pred. No. 7.7e-36;
  Matches  235;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy     139 GGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAG 198
           |||||   ||  || |||| ||||| |     |||| |  |||  | |   || ||||
Db   16537 GGAGGCAGCCTCCGCGGAGGAGCAGAAGGAAAAGGAAGATAAAGTGACTAGTCCAGAGAA 16596

Qy     199 AGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTC 258
           ||||||||| |||||  |||| || |  || ||     || ||||||||||||||||| ||
Db   16597 AGCTGAAGAAGCAAAATTAAAAGCAAGGTATCCTCATCTGGGACAAAAGCCTGGAGGTTC 16656

Qy     259 CGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAA 318
           || ||| | | |||  || | ||||||||||||||| || ||||| || ||  | |||||
Db   16657 AGATTTCTTAAGGAAACGATTGCAGAAAGGGCAAAAATATTTTGATTCTGGGCATTACAA 16716

Qy     319 CATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCT 378
           || ||| ||||| || | |||||  ||||| || || | |||||  ||||| ||||
Db   16717 CAGGGCTAAAGCAAAAACGAAGAGCAAGCAACTTCCTACTGCAGCTCCAGATAAGACGGA 16776

Qy     379 GGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGT 438
           ||| |||||||| ||||| || || |||||||| || || ||  | ||  | || || ||
Db   16777 GGTCACTGGTGACCACATTCCTACTCCACAGGACCTTCCTCAACGGAAACCATCCCTTGT 16836

Qy     439 CACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCT 486
             | | |||||| || ||  |   |   ||      |||| |  |||
Db   16837 TGCTAACAAGCTGGCTGGCTGATTAAAAGAGCTGAACTGCATGAAGCT 16884


RESULT  34
SSA005981
LOCUS       SSA005981              339 bp    mRNA    linear   MAM 07-JAN-1999
DEFINITION  Sus scrofa mRNA for cAMP-regulated phosphoprotein (ARPP-19).
ACCESSION   AJ005981
VERSION     AJ005981.1  GI:4138554
KEYWORDS    ARPP-19 gene; cAMP-regulated phosphoprotein.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Bataille D., U 376 'Endocrinologie des
            Peptides et regulation genique', INSERM, CHU Arnaud-de-Villeneuve,
            34295, FRANCE
REFERENCE   2  (bases 1 to 339)
  AUTHORS   Girault,J.A., Horiuchi,A., Gustafson,E.L., Rosen,N.L. and
            Greengard,P.
  TITLE     Differential expression of ARPP-16 and ARPP-19, two highly related
            cAMP-regulated phosphoproteins, one of which is specifically
            associated with dopamine-innervated brain regions
  JOURNAL   J. Neurosci. 10 (4), 1124-1133 (1990)
  MEDLINE   90229913
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
     gene            1. .339
                     /gene="ARPP-19"
     CDS             1. .339
                     /gene="ARPP-19"
                     /codon_start=1
                     /product="cAMP-regulated phosphoprotein"
                     /protein_id="CAA06795.1"
                     /db_xref="GI:4138555"
                     /translation="MSAEVPEAASAEEQKEMEDKVTSPEKAEEAKLKARYPHLGQKPG
                     GSDFLRKRLQKGQKYFDSGDYNMAKAKMKNKQLPTAAPDKTEVTGDHIPTPQDLPQRK
                     PSLVASKLAG"
```

```
BASE COUNT      115 a     74 c     84 g     66 t
ORIGIN


  Query Match             13.9%;  Score 167;  DB 4;  Length 339;
  Best Local Similarity   70.2%;  Pred. No. 3e-36;
  Matches  224;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy    138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
          | ||||   ||  || |||| ||||| |      |||  |  |||  | |   || ||||
Db     17 CCGAGGCAGCCTCCGCGGAGGAGCAGAAGGAAATGGAAGATAAAGTGACTAGTCCAGAGA 76

Qy    198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
          |||||||||  |||||  |||| || |  || ||    || ||||||||||||||||| |
Db     77 AAGCTGAAGAAGCAAAATTAAAAGCGAGGTATCCTCATCTGGGACAAAAGCCTGGAGGTT 136

Qy    258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
          | || ||| | |  ||  || | |||||||||||||| || ||||| || || || || |
Db    137 CAGATTTCTTAAGAAAACGATTGCAGAAAGGGCAAAAATATTTTGATTCTGGGGATTATA 196

Qy    318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
          ||||||| ||||| || |||||||| |||||  || || | |||||  || || ||||
Db    197 ACATGGCTAAAGCAAAAATGAAGAACAAGCAACTTCCTACTGCAGCCCCGGATAAGACAG 256

Qy    378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
          |||  |||||||| |||||  |||||  |||||||| || || ||  | ||  | || || |
Db    257 AGGTCACTGGTGACCACATTCCCACTCCACAGGACCTTCCTCAACGGAAACCATCTCTTG 316

Qy    438 TCACCAGCAAGCTTGCGGG 456
          |  | |||||  || || ||
Db    317 TTGCTAGCAAACTGGCTGG 335


RESULT  35
AK057673
LOCUS       AK057673               2288 bp    mRNA    linear   PRI 31-OCT-2001
DEFINITION  Homo sapiens cDNA FLJ33111 fis, clone TRACH2001085.
ACCESSION   AK057673
VERSION     AK057673.1  GI:16553438
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens trachea cDNA to mRNA, clone_lib:TRACH2
            clone:TRACH2001085.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
            Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J.,
            Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K.,
            Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H.,
            Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A.,
            Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.
            and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2288)
  AUTHORS   Isogai,T., Otsuki,T. and Sugiyama,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: RAB and
            HRI.
FEATURES             Location/Qualifiers
     source          1. .2288
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="TRACH2001085"
                     /tissue_type="trachea"
                     /clone_lib="TRACH2"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      604 a    462 c    597 g    625 t
ORIGIN


  Query Match             13.8%;  Score 165.4;  DB 9;  Length 2288;
  Best Local Similarity   96.6%;  Pred. No. 1.2e-35;
  Matches  169;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     22 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 81

Qy     61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     82 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 141

Qy    121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAG 175
          ||||||||||||||||||||||||||||||||||||||||||||  | |  | ||
Db    142 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGTGAAGTGGAAG 196


RESULT  36
BOVARPP16
LOCUS       BOVARPP16               1961 bp    mRNA    linear   MAM 26-APR-1993
DEFINITION  Bovine cAMP-regulated phosphoprotein (ARPP-16) mRNA, complete cds.
ACCESSION   M33617 J05450
VERSION     M33617.1  GI:162687
KEYWORDS    ARPP-16; cAMP-regulated phosphoprotein.
SOURCE      Bovine young adult caudate nuclei, cDNA to mRNA, clone 6.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1961)
  AUTHORS   Horiuchi,A., Williams,K.R., Kurihara,T., Nairn,A.C. and
            Greengard,P.
  TITLE     Purification and cDNA cloning of ARPP-16, a cAMP-regulated
            phosphoprotein enriched in basal ganglia, and of a related
            phosphoprotein, ARPP-19
  JOURNAL   J. Biol. Chem. 265, 9476-9484 (1990)
  MEDLINE   90264450
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by A.Horiuchi, 06-APR-1990.
FEATURES             Location/Qualifiers
     source          1. .1961
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
     mRNA            1. .1961
                     /product="ARPP-16 mRNA"
     misc_feature    85. .88
                     /note="phosphorylation site"
     CDS             112. .402
                     /note="cAMP-regulated phosphoprotein (ARPP-19)"
                     /codon_start=1
                     /protein_id="AAA30385.1"
                     /db_xref="GI:162688"
                     /translation="MEDKVTSPEKAEEAKLKARYPHLGQKPGGSDFLRKRLQKGQKYF
                     DSGDYNMAKAKMKNKQLPTATPDKTEVTGDHIPTPQDLPQRKPSLVASKLAG"
BASE COUNT      606 a    320 c    355 g    680 t
ORIGIN


  Query Match             13.2%;  Score 158.6;  DB 4;  Length 1961;
  Best Local Similarity   72.3%;  Pred. No. 1e-33;
  Matches  206;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy    172 GGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCC 231
          |||  |  |||  | |   || |||| |||||||||| |||||  |||| || |  || ||
```

```
Db    114 GGAAGATAAAGTGACTAGTCCAGAGAAAGCTGAAGAAGCAAAATTAAAAGCAAGGTATCC 173

Qy    232 AAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCA 291
          || ||||||||||||||||| ||||| ||  | | |||  || | ||||||||||||
Db    174 TCATCTGGGACAAAAGCCTGGAGGTTCCGATTTTTTAAGGAAACGATTGCAGAAAGGGCA 233

Qy    292 AAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCT 351
          ||| || ||||| || || || |||||||||| ||||| || |||||||| ||||| ||
Db    234 AAAATATTTTGATTCTGGGGATTACAACATGGCTAAAGCAAAAATGAAGAACAAGCAACT 293

Qy    352 GCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGA 411
          || | ||||   || || ||||    ||| |||||||| ||||| ||||| ||||||||
Db    294 TCCTACTGCAACCCCGGATAAGACAGAGGTCACTGGTGACCACATTCCCACTCCACAGGA 353

Qy    412 TCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
          || || ||  | ||  | || || ||  | ||||| || || ||
Db    354 CCTTCCTCAACGGAAACCATCTCTTGTTGCTAGCAAACTGGCTGG 398


RESULT  37
AC068538
LOCUS       AC068538               173169 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens BAC clone RP11-141G4 from 2, complete sequence.
ACCESSION   AC068538
VERSION     AC068538.8  GI:16931042
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173169)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 173169)
  AUTHORS   VanBrunt,A., Haglund,K. and Doebber,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-141G4
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 173169)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 173169)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 173169)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 173169)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Nov 15, 2001 this sequence version replaced gi:16418235.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0141G04
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-292K15; the clone sequenced
            to the right is RP11-480J15.  Actual start of this clone is at base
            position 1 of RP11-141G4; actual end is at base position 173169 of
            RP11-141G4.

            Data from AC092845 was used to finish this clone, AC068538.
            Polymorphisms have been identified between AC092845 and AC068538.
FEATURES             Location/Qualifiers
     source          1. .173169
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-141G4"
                     /clone_lib="RPCI-11"
     repeat_region   2. .803
                     /rpt_family="ERV1"
     repeat_region   812. .909
                     /rpt_family="ERV1"
     repeat_region   910. .1279
                     /rpt_family="ERV1"
     repeat_region   2183. .2217
                     /rpt_family="ERV1"
     repeat_region   2220. .2653
                     /rpt_family="ERV1"
     repeat_region   2654. .3050
                     /rpt_family="MaLR"
     repeat_region   3057. .3166
                     /rpt_family="ERV1"
     repeat_region   3639. .3981
                     /rpt_family="L1"
     repeat_region   4152. .4444
                     /rpt_family="Alu"
     repeat_region   4607. .5058
                     /rpt_family="MaLR"
     repeat_region   5362. .5449
                     /rpt_family="MaLR"
     repeat_region   5552. .5840
                     /rpt_family="Alu"
     repeat_region   5869. .5968
                     /rpt_family="L1"
     repeat_region   5955. .6018
                     /rpt_family="CT-rich"
     repeat_region   6379. .6783
```

```
                /rpt_family="ERVL"
repeat_region   6782. .6802
                /rpt_family="AT_rich"
repeat_region   6806. .6837
                /rpt_family="(TG)n"
repeat_region   6899. .7204
                /rpt_family="Alu"
repeat_region   7208. .7716
                /rpt_family="MER2_type"
repeat_region   7716. .7807
                /rpt_family="MER2_type"
repeat_region   8105. .8353
                /rpt_family="L1"
repeat_region   8610. .8980
                /rpt_family="MaLR"
repeat_region   9039. .9059
                /rpt_family="AT_rich"
repeat_region   9782. .10081
                /rpt_family="Alu"
repeat_region   10260. .10487
                /rpt_family="MIR"
repeat_region   12061. .12147
                /rpt_family="CR1"
repeat_region   12403. .12646
                /rpt_family="CR1"
repeat_region   12645. .12983
                /rpt_family="CR1"
repeat_region   13017. .13087
                /rpt_family="CR1"
repeat_region   13831. .14143
                /rpt_family="L1"
repeat_region   14231. .14268
                /rpt_family="(TA)n"
repeat_region   14583. .14867
                /rpt_family="Alu"
repeat_region   15161. .15391
                /rpt_family="MIR"
repeat_region   15906. .16029
                /rpt_family="MIR"
repeat_region   16219. .16364
                /rpt_family="L2"
repeat_region   16401. .16709
                /rpt_family="Alu"
repeat_region   16683. .16714
                /rpt_family="(GAAAA)n"
repeat_region   18661. .18943
                /rpt_family="Alu"
repeat_region   19142. .19417
                /rpt_family="Alu"
repeat_region   20259. .20316
                /rpt_family="MIR"
repeat_region   20650. .20746
                /rpt_family="AcHobo"
repeat_region   21564. .21733
                /rpt_family="MER1_type"
repeat_region   22242. .22348
                /rpt_family="MIR"
repeat_region   22408. .22556
                /rpt_family="MER103"
repeat_region   24087. .24364
                /rpt_family="ERV1"
repeat_region   24366. .24949
                /rpt_family="ERV1"
repeat_region   24950. .25377
                /rpt_family="ERV1"
repeat_region   25576. .25811
                /rpt_family="MIR"
repeat_region   25935. .26516
                /rpt_family="ERV1"
repeat_region   26517. .26825
                /rpt_family="Alu"
repeat_region   26826. .26973
                /rpt_family="ERV1"
repeat_region   28124. .28315
                /rpt_family="MIR"

  Query Match             13.1%;  Score 157.2;  DB 9;  Length 173169;
  Best Local Similarity   68.0%;  Pred. No. 6.4e-33;
  Matches  219;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy    138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
          | ||||   ||   | |||| ||||| |       |||  |   ||  | |   || ||||
Db  49976 CCGAGGCAGCCTCTGTGGAGGAGCAGAAGGAAATGGAAGATACAGTGACTAGTCCAGAGA 50035

Qy    198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
           || |||||  |||||  |||| || | ||| ||    || |||||||| ||||||||| |
Db  50036 AAGTTGAAGCAGCAAAATTAAAAGCAAGATATCCTCATCTGGGACAAATGCCTGGAGGTT 50095

Qy    258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
          | || ||| | | |||  || | ||||||||||||||| || ||||| || || || ||||
Db  50096 CAGATTTCTTAAGGAAACGATTGCAGAAAGGGCAAAAATATTTTGATTCTGGGGATTACA 50155

Qy    318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
          ||||||| ||||| || |||||||| |||| || || | |||||  |  || ||||
Db  50156 ACATGGCTAAAGCAAAAATGAAGAACGAGCAACTTCCTACTGCAGCTCTGGATAAGATGG 50215

Qy    378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
           ||| || ||||| ||||| ||||| ||| |||| || || ||    ||  | ||  | |
Db  50216 AGGTCACGGGTGACCACATTCCCACTCCAGAGGACCTTCCTCAACAGAAACCATCCATTG 50275

Qy    438 TCACCAGCAAGCTTGCGGGTGG 459
          |  | |||||||| || ||| |
Db  50276 TTGCTAGCAAGCTGGCTGGTTG 50297


RESULT  38
AC074041/c
LOCUS       AC074041              170548 bp    DNA     linear   ROD 28-JUN-2001
DEFINITION  Mus musculus chromosome 2 clone RP23-258D16 strain C57BL6/J,
            complete sequence.
ACCESSION   AC074041
VERSION     AC074041.3  GI:14573692
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 170548)
  AUTHORS   Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L.,
            Granite,S., Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E.,
            Lee-Lin,S.-Q., Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B.,
            Masiello,C., Mastrian,S.D., McCloskey,J.C., McDowell,J.,
            Pearson,R., Prasad,A., Shevchenko,Y., Snyder,B., Stantripop,S.,
            Thomas,J.W., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L.,
            Walker,M.A., Wetherby,K.D., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170548)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2000) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 170548)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jun 28, 2001 this sequence version replaced gi:11024833.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
                 Center project name: ta
```

```
          Center clone name: 258D16

          This sequence was finished as follows unless otherwise noted:
          all regions were double-stranded, sequenced with an
          alternate chemistry, or covered by high quality data
          (i.e., phred quality >= 30); an attempt was made to resolve
          all sequencing problems, such as compressions and repeats;
          all regions were covered by at least one plasmid subclone
          or more than one M13 subclone; and the assembly was confirmed
          by restriction digest.

          CLONE LENGTH: This sequence represents the entire insert of
          this clone unless otherwise noted. If there are overlapping
          clones, the overlaps are noted in the beginning and end of
          the Features section.
FEATURES             Location/Qualifiers
     source          1. .170548
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-258D16"
                     /clone_lib="RPCI mouse BAC library 23"
     misc_feature    1. .17547
                     /note="clone overlaps with GenBank Accession Number
                     AC074224 (nucleotides 171443-188989) clone RP23-43C3
                     (center project name uz)"
     misc_feature    170267. .170274
                     /note="low quality single stranded/single chemistry
                     region"
BASE COUNT    51882 a  37418 c  35891 g  45357 t
ORIGIN


  Query Match             12.4%;  Score 148.8;  DB 10;  Length 170548;
  Best Local Similarity   59.6%;  Pred. No. 1.6e-30;
  Matches  567;  Conservative    0;  Mismatches  292;  Indels   93;  Gaps   15;

Qy     120 AAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAG 179
           || |||| ||||| |  || ||||||| | | |  ||||     ||||| | ||| | ||
Db  120869 AACAAGAGGAGAAGCAAGCAGAGGAGATCTGTGCAGAGA-----GACACTCGGGAAAGAG 120815

Qy     180 AAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAG 239
           ||| | |  | ||| |||||||     |||||||||  || |||  ||||  |||||||
Db  120814 AAGCTCTCTTCCCTAAGAGAGCAAGGAAGGCAAAGCTGTAGACCAGGTACCTGAGCCTAG 120755

Qy     240 GACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACT 299
           |||  || |||||||  | |   ||       |||| ||| ||||| || || || | |
Db  120754 CACAGGAGACTGGAGGTGCTG--CTCTCTCATGAGAGGCTCTAGAAAAGGAAAGAG-ATG 120698

Qy     300 TTGACTCAGGAGAC----TACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCA 355
           || ||||||||||     | | | ||| || | || |||||||||| | | |  ||||
Db  120697 TTTACTCAGGAGACGAGAGAGAGCGTGGTCAGAACCGAGATGAAGAACAGGGAAATGCC- 120637

Qy     356 AGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTG 415
                       || |  ||     || | ||||||| |||||||| || ||
Db  120638 ------------------ACATCCTGGTAACTGCTAACATCCCTACCCCACAAGACCTT 120598

Qy     416 CCCCAGAGAAAGTCCTCGCT----CGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATG 471
           |||||  |||||||||| ||     |||||||||||||||| |||||||||| |||| |
Db  120597 CCCCATTGAAAGTCCTCACTTGTCTGTCACCAGCAAGCTTGCTGGTGGCCAAGATGAAGG 120538

Qy     472 A------------TGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCC 519
           |            |||||   |||| || ||    |||         | |  || |   |
Db  120537 ATGCTATGTGGGGTGCTGTTTGGGGTTCAGCAGTGTCCACCTGGGTCCTAGGCTGTGGCT 120478

Qy     520 CCTCCCTGCCCCACCCGGGTCCT--GTGCTGGCTCCTGCCCCTTCAGCCAGGGGTCAGGA 577
           ||| ||   ||  |  ||   ||  | |    |   || | |    |  |||     ||
Db  120477 CCTACCCTACCTGCTTGGAGGCTAAGAGTCAGAAAATGTCTCAGGTGTGAGGCACGGGGC 120418

Qy     578 GGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATG 637
           | ||   ||  | | |  |  | ||||| | ||  | ||   ||   | |||||| |||
Db  120417 TGGGGTGGGGTGGGGTGGGGTGGGGGCAGCAACCTCTGCCCTATGACATTCCAGCAGATG 120358

Qy     638 GAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTC----- 692
           | | | | ||||||||||||||||    ||||||  || |  | ||    |||| |
Db  120357 G-GTCACCTGGGCTGAGCACCAAAGTTTTGAACTATTATATTGTTTAACTTTTCCTCCCC 120299

Qy     693 --CAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAG 750
             |||||||| ||||||| |||| |||||||||| |      ||    |  || | | ||
Db  120298 CTCAAATAACTGTTGGGATAAATTTCAATGAAATCCC-----TGTCTATTAGGTTAAGAG 120244

Qy     751 GGTGGGGTGGGAGATATGGAGGAGTATGAATTAGGGCTTGGAGTTGGTAAAAA------- 803
           || | |||||||| | |||    |  ||||   ||||  || | ||| | || |
Db  120243 GGCGTGGTGGGAGGTTTGGGTCAAAATGAGCGAGGGTCTGCAATTGATTAAGACCTAAAC 120184

Qy     804 ------------CATTCCTGACTATCCTTCTTAACCACGTGGCTGATGTGGGGTAGGTAT 851
                       |||||  ||   |||| |||  |  |||||||| || |  | |||||
Db  120183 AAACAAAATAGCCATTCAAGATGGTCCTCCTTGCCTGCGTGGCTGAGGTAGTATGGGTAT 120124

Qy     852 GAG------GGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTAGTTGAGCTCTG-- 903
           |       ||| |||   ||   ||| ||||||| |||| ||||||| | ||| |||
Db  120123 AAAAGGGGTGGGGAGGGTTTGTCTTAGACTAATGACAAGGAGTTCTAGCTCAGCCCTGAA 120064

Qy     904 TAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATAAAAGAAACTCCAT 963
           |||  || || ||| ||   |||| ||  |||||| | ||||| || |||| || | |
Db  120063 AAGAAGAAAGCATTGGTTGGATGTGTCTGCGGACCTGTTATCAGAGAAGAGAAGCTGCTG 120004

Qy     964 CCGCACAGACA---GATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAGTTGA 1012
               || |||   ||| ||| ||||||||  || |||| ||||| || | |
Db  120003 ATCTCCACACAGTTGATACAAGCAGCTCCTTCAGCTCTGTAGAGCAAGCTCA 119952


RESULT  39
AC090836/c
LOCUS       AC090836               70244 bp    DNA     linear   HTG 10-MAR-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-23N2 map 15, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC090836
VERSION     AC090836.1  GI:13270705
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 70244)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-23N2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 70244)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
```

```
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
    Center project name: L13121
    Center clone name: 23_N_2
--------------.
* NOTE: This record contains 87 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*       1      703: contig of 703 bp in length
*     704 803: gap of      100 bp
*     804     1525: contig of 722 bp in length
*    1526 1625: gap of      100 bp
*    1626     2347: contig of 722 bp in length
*    2348 2447: gap of      100 bp
*    2448     3133: contig of 686 bp in length
*    3134 3233: gap of      100 bp
*    3234     3943: contig of 710 bp in length
*    3944 4043: gap of      100 bp
*    4044     4754: contig of 711 bp in length
*    4755 4854: gap of      100 bp
*    4855     5563: contig of 709 bp in length
*    5564 5663: gap of      100 bp
*    5664     6380: contig of 717 bp in length
*    6381 6480: gap of      100 bp
*    6481     7207: contig of 727 bp in length
*    7208 7307: gap of      100 bp
*    7308     8022: contig of 715 bp in length
*    8023 8122: gap of      100 bp
*    8123     8826: contig of 704 bp in length
*    8827 8926: gap of      100 bp
*    8927     9602: contig of 676 bp in length
*    9603 9702: gap of      100 bp
*    9703    10410: contig of 708 bp in length
*   10411 10510: gap of      100 bp
*   10511    11209: contig of 699 bp in length
*   11210 11309: gap of      100 bp
*   11310    12023: contig of 714 bp in length
*   12024 12123: gap of      100 bp
*   12124    12827: contig of 704 bp in length
*   12828 12927: gap of      100 bp
*   12928    13638: contig of 711 bp in length
*   13639 13738: gap of      100 bp
*   13739    14445: contig of 707 bp in length
*   14446 14545: gap of      100 bp
*   14546    15212: contig of 667 bp in length
*   15213 15312: gap of      100 bp
*   15313    16027: contig of 715 bp in length
*   16028 16127: gap of      100 bp
*   16128    16828: contig of 701 bp in length
*   16829 16928: gap of      100 bp
*   16929    17634: contig of 706 bp in length
*   17635 17734: gap of      100 bp
*   17735    18453: contig of 719 bp in length
*   18454 18553: gap of      100 bp
*   18554    19272: contig of 719 bp in length
*   19273 19372: gap of      100 bp
*   19373    20047: contig of 675 bp in length
*   20048 20147: gap of      100 bp
*   20148    20861: contig of 714 bp in length
*   20862 20961: gap of      100 bp
*   20962    21670: contig of 709 bp in length
*   21671 21770: gap of      100 bp
*   21771    22479: contig of 709 bp in length
*   22480 22579: gap of      100 bp
*   22580    23299: contig of 720 bp in length
*   23300 23399: gap of      100 bp
*   23400    24127: contig of 728 bp in length
*   24128 24227: gap of      100 bp
*   24228    24937: contig of 710 bp in length
*   24938 25037: gap of      100 bp
*   25038    25750: contig of 713 bp in length
*   25751 25850: gap of      100 bp
*   25851    26524: contig of 674 bp in length
*   26525 26624: gap of      100 bp
*   26625    27338: contig of 714 bp in length
*   27339 27438: gap of      100 bp
*   27439    28163: contig of 725 bp in length
*   28164 28263: gap of      100 bp
*   28264    28988: contig of 725 bp in length
*   28989 29088: gap of      100 bp
*   29089    29801: contig of 713 bp in length
*   29802 29901: gap of      100 bp
*   29902    30602: contig of 701 bp in length
*   30603 30702: gap of      100 bp
*   30703    31412: contig of 710 bp in length
*   31413 31512: gap of      100 bp
*   31513    32238: contig of 726 bp in length
*   32239 32338: gap of      100 bp
*   32339    33068: contig of 730 bp in length
*   33069 33168: gap of      100 bp
*   33169    33887: contig of 719 bp in length
*   33888 33987: gap of      100 bp
*   33988    34697: contig of 710 bp in length
*   34698 34797: gap of      100 bp
*   34798    35508: contig of 711 bp in length
*   35509 35608: gap of      100 bp
*   35609    36320: contig of 712 bp in length
*   36321 36420: gap of      100 bp
*   36421    37113: contig of 693 bp in length
*   37114 37213: gap of      100 bp
*   37214    37934: contig of 721 bp in length
*   37935 38034: gap of      100 bp
*   38035    38725: contig of 691 bp in length
*   38726 38825: gap of      100 bp
*   38826    39537: contig of 712 bp in length
*   39538 39637: gap of      100 bp
*   39638    40342: contig of 705 bp in length
*   40343 40442: gap of      100 bp
*   40443    41151: contig of 709 bp in length
*   41152 41251: gap of      100 bp
*   41252    41966: contig of 715 bp in length
*   41967 42066: gap of      100 bp
*   42067    42774: contig of 708 bp in length
*   42775 42874: gap of      100 bp
*   42875    43580: contig of 706 bp in length
*   43581 43680: gap of      100 bp
*   43681    44360: contig of 680 bp in length
*   44361 44460: gap of      100 bp
*   44461    45162: contig of 702 bp in length
*   45163 45262: gap of      100 bp
*   45263    45957: contig of 695 bp in length
*   45958 46057: gap of      100 bp
*   46058    46774: contig of 717 bp in length
*   46775 46874: gap of      100 bp
*   46875    47570: contig of 696 bp in length
*   47571 47670: gap of      100 bp
*   47671    48380: contig of 710 bp in length
*   48381 48480: gap of      100 bp
*   48481    49195: contig of 715 bp in length
*   49196 49295: gap of      100 bp
*   49296    49998: contig of 703 bp in length
*   49999 50098: gap of      100 bp
*   50099    50820: contig of 722 bp in length
```

```
            *    50821 50920: gap of      100 bp
            *    50921    51624: contig of 704 bp in length
            *    51625 51724: gap of      100 bp
            *    51725    52433: contig of 709 bp in length
            *    52434 52533: gap of      100 bp
            *    52534    53228: contig of 695 bp in length
            *    53229 53328: gap of      100 bp
            *    53329    54016: contig of 688 bp in length
            *    54017 54116: gap of      100 bp
            *    54117    54835: contig of 719 bp in length
            *    54836 54935: gap of      100 bp
            *    54936    55651: contig of 716 bp in length
            *    55652 55751: gap of      100 bp

  Query Match              8.3%;  Score 99.2;  DB 2;  Length 70244;
  Best Local Similarity   71.2%;  Pred. No. 1.6e-16;
  Matches  131;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy     289 GCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCA 348
           |||||| || ||||| || || || ||||||||||| ||||| || |||||||| |||||
Db   20733 GCAAAAATATTTTGATTCTGGGGATTACAACATGGCTAAAGCAAAAATGAAGAACAAGCA 20674

Qy     349 GCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACA 408
           || || | |||||  || || ||||    ||| |||||||| ||||| ||||| || ||
Db   20673 ACTTCCTACTGCAGCTCCGGATAAGACGGAGGTCACTGGTGACCACATTCCCACTCCGCA 20614

Qy     409 GGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGA 468
           || || || ||  | ||| | || || ||  | |||||||| || ||  |   |   |
Db   20613 AGACCTTCCTCAACGGAAGCCGTCCCTTGTTGCTAGCAAGCTGGCTGGCTGATTAAGAGG 20554

Qy     469 ATGA 472
           |||
Db   20553 CTGA 20550


RESULT  40
AC025917/c
LOCUS       AC025917               149081 bp    DNA     linear   HTG 17-JAN-2002
DEFINITION  Homo sapiens chromosome 15 clone RP11-23N2 map 15q21, ***
            SEQUENCING IN PROGRESS ***, 1 ordered pieces.
ACCESSION   AC025917
VERSION     AC025917.7  GI:18182752
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 149081)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Sequencing of human chromosome 15 D15S146-D15S117 region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 149081)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
            James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
            and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
COMMENT     On Jan 17, 2002 this sequence version replaced gi:13194215.
            ------------ Genome Center
                Center: Multimegabase Sequencing Center
                Center code: UWMSC
                Web site: http://chroma.mbt.washington.edu/msg_www
                Contact: leerowen@systemsbiology.org
            ------------ Summary Statistics
                Sequencing vector: pUC18; L08752
                Chemistry: Dye-terminator Big Dye; 90% of reads
                Chemistry: Dye-primer Big Dye; 10% of reads
                Assembly program: Phrap; version 0.990399
                Insert size: 142000; agarose-fp
                Quality coverage: 12.8x in Q20 bases; sum-of-contigs
            ------------.
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   149081: contig of 149081 bp in length.
FEATURES            Location/Qualifiers
     source         1. .149081
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="15"
                    /map="15q21"
                    /clone="RP11-23N2"
                    /clone_lib="RPCI Human BAC library 11"
                    /note="This clone overlaps RP11-327C2 and RP11-519C12."
BASE COUNT    46006 a  30622 c  29456 g  42997 t
ORIGIN


  Query Match              8.3%;  Score 99.2;  DB 2;  Length 149081;
  Best Local Similarity   71.2%;  Pred. No. 1.9e-16;
  Matches  131;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy     289 GCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCA 348
           |||||| || ||||| || || || ||||||||||| ||||| || |||||||| |||||
Db   96516 GCAAAAATATTTTGATTCTGGGGATTACAACATGGCTAAAGCAAAAATGAAGAACAAGCA 96457

Qy     349 GCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACA 408
           || || | |||||  || || ||||    ||| |||||||| ||||| ||||| || ||
Db   96456 ACTTCCTACTGCAGCTCCGGATAAGACGGAGGTCACTGGTGACCACATTCCCACTCCGCA 96397

Qy     409 GGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGA 468
           || || || ||  | ||| | || || ||  | |||||||| || ||  |   |   |
Db   96396 AGACCTTCCTCAACGGAAGCCGTCCCTTGTTGCTAGCAAGCTGGCTGGCTGATTAAGAGG 96337

Qy     469 ATGA 472
           |||
Db   96336 CTGA 96333


Search completed: September 12, 2002, 15:47:42
Job time: 21926 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2002, 16:14:20 ; Search time 407.7 Seconds
(without alignments)
5049.253 Million cell updates/sec

Title: US-09-824-178-2
Perfect score: 1199
Sequence: 1 GTGACAGGAGCCGAAGCAGC.....CCAAAAAAAAAAAAAAAAAA 1199

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1199	100.0	1199	19	AAV45211	Human endosulfine
2	1131.4	94.4	1208	22	AAH14942	Human cDNA sequenc
3	570.8	47.6	808	22	AAH07746	Human cDNA clone (
c 4	515.6	43.0	587	22	AAH12719	Human cDNA clone (
5	503	42.0	553	24	ABK15891	Human lung tumour
6	501.8	41.9	669	20	AAZ24516	Human lung tumor a
7	501.8	41.9	669	21	AAC65755	Human lung cancer-
8	461.2	38.5	503	21	AAC01495	Human secreted pro
9	456	38.0	1103	19	AAV45210	Human endosulfine
10	452	37.7	494	24	ABK15892	Human lung tumour
11	420.8	35.1	863	21	AAC98847	Human pancreatic c
12	346.4	28.9	454	19	AAV30285	Human phosphoprote
13	300.6	25.1	312	21	AAC06352	Human secreted pro
c 14	285.8	23.8	314	14	AAQ60964	Human brain Expres
15	283	23.6	343	14	AAQ60963	Human brain Expres
16	168.6	14.1	2498	22	AAI93443	Human polynucleoti
c 17	165.4	13.8	2544	22	AAK90275	Human digestive sy
c 18	165.4	13.8	2544	22	AAI57655	Human colorectal c
19	155.4	13.0	3316	19	AAV38840	Homo sapiens C4-2
20	121.8	10.2	341	21	AAC00778	Human secreted pro
21	121.8	10.2	350	21	AAC00776	Human secreted pro
22	114	9.5	295	21	AAC00775	Human secreted pro
23	111.8	9.3	407	21	AAC00777	Human secreted pro
24	67.8	5.7	91	21	AAC15418	Human secreted pro
25	47.6	4.0	1835	20	AAX27245	Human CLAR1 coding
26	47.4	4.0	551	21	AAZ52551	Human secreted pro
27	47	3.9	1133	22	AAS01015	Maize disease resi
28	46.2	3.9	1613	20	AAX26196	DNA sequence of gl
29	45.6	3.8	405	22	AAI84623	Human polynucleoti
30	45.6	3.8	453	22	AAI84652	Human polynucleoti
31	45.6	3.8	840	21	AAC78104	Human cancer assoc
32	45.6	3.8	2968	22	AAF32706	Human secreted pro
c 33	44.8	3.7	794	22	AAI95102	Human neuroblastom
34	44.8	3.7	2115	22	AAL26422	Human breast cance
35	44.6	3.7	400	22	AAI88751	Human polynucleoti
36	44.6	3.7	422	22	AAI91888	Human polynucleoti
c 37	44.6	3.7	438	22	AAI93088	Human polynucleoti
c 38	44.6	3.7	457	22	AAL19946	Human breast cance
39	44.4	3.7	449	22	AAI84684	Human polynucleoti
40	44.4	3.7	2262	21	AAZ52560	Human secreted pro
41	44.2	3.7	403	22	AAI82937	Human polynucleoti
42	44.2	3.7	411	22	AAI92340	Human polynucleoti
43	44	3.7	2968	21	AAC78190	Human cancer assoc
44	44	3.7	2968	22	AAH34408	Human colon cancer
45	43.8	3.7	389	22	AAS29135	cDNA encoding for

ALIGNMENTS

RESULT 1
AAV45211
ID AAV45211 standard; cDNA; 1199 BP.
XX
AC AAV45211;
XX
DT 19-OCT-1998 (first entry)
XX
DE Human endosulfine B gene.
XX
KW ds; endosulfine; central nervous system; pancreatic beta-cell; pain;
KW insulin; cytoprotective agent; diabetes; epilepsy; depression; ischemia.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 107..472
FT /*tag= a
FT /product= "Endosulfine B protein"
XX
PN WO9830692-A2.
XX
PD 16-JUL-1998.
XX
PF 07-JAN-1998; 98WO-US00137.
XX
PR 07-JAN-1997; 97US-0779775.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Anderson KL, Roch J, Scott VE, Sullivan JP;
XX

```
DR   WPI; 1998-399140/34.
DR   P-PSDB; AAW61552.
XX
PT   Novel human endosulfine A and B - for screening for compounds
PT   modulating endosulfine receptor activity, useful as cytoprotective
PT   agent(s)
XX
PS   Claim 2; Page 58; 76pp; English.
XX
CC   The human endosulfine A and B genes encode proteins related to other
CC   members of the endosulfine family which have been shown to bind receptors
CC   from both the central nervous system and pancreatic beta-cells and to
CC   induce insulin secretion from a rodent beta-cell tumour line.  The
CC   protein can be used to screen for compounds that modulate endosulfine
CC   receptor activity, especially as a cytoprotective agent.
CC   Modulators of endosulfine activity may represent a treatment of
CC   diabetes, epilepsy, pain, depression and ischemia.
XX
SQ   Sequence 1199 BP; 335 A; 288 C; 333 G; 243 T; 0 other;


  Query Match             100.0%;  Score 1199;  DB 19;  Length 1199;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1199;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gtgacaggagccgaagcagcagcgcaggttgtccccgtttccctccccttccttctc 60

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 cggttgccttcccgggcccccttacactccacagtcccggtcccgccatgtcccagaaaca 120

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 agaagaagagaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaaga 180

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aggtattctgcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctagg 240

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 acaaaagcctggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactt 300

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 tgactcaggagactacaacatggccaaagccaagatgaagaataagcagctgccaagtgc 360

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 aggaccagacaagaacctggtgactggtgatcacatccccaccccacaggatctgcccca 420

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 gagaaagtcctcgctcgtcaccagcaagcttgcgggtggccaagttgaatgatgctgccc 480

Qy     481 GGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ggggctctgccagatcctgagactgcttttgccgcttccccteccctgccccacccgggtc 540

Qy     541 CTGTGCTGGCTCCTGCCCCTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ctgtgctggctcctgcccctcagccaggggtcaggaggtggctcgggtgtgggctggag 600

Qy     601 AGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 aggcagaagcccttcctgttggtgtcccagcacatggagcccctgggctgagcaccaa 660

Qy     661 GACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAATG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 gaccttgaacctttttgttttacctttttccaaataacagttgggagaaatatcaatg 720

Qy     721 AAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGGAGGAGTATGAA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 aaattctgggggtgggggtggggctgaaagggtggggtgggagatatggaggagtatgaa 780

Qy     781 TTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACGTGGCTGATGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 ttagggcttggagttggtaaaaacattcctgactatccttcttaaccacgtggctgatgt 840

Qy     841 GGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTAGTTGAGCT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ggggtaggtatgaggggaaggaagtggagtagcctaatgaaaaggggttctagttgagct 900

Qy     901 CTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATAAAAGAAACTC 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 ctgtagataaatgccttgtttcagtgtggttggagacctggtgtcagataaaagaaactc 960

Qy     961 CATCCGCACAGACAGATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAGTTGAGAACTCAA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 catccgcacagacagatgcaaacagctcctctagttctgcagagctagttgagaactcaa 1020

Qy    1021 CATTAATCATTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGGGAAGAAGGGCAGT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 cattaatcattttaaaaagtactgtccttgaaatagatttgctgtgggaagaagggcagt 1080

Qy    1081 GAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTTC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 gagtgtgggagaaaggagccgtgagcgtggggaaccccacagagcccaaaggacttttttc 1140

Qy    1141 AGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAAAAAAAAA 1199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 agtattcgaaataaacaaaacaaaaacccatgaaaaaacccaaaaaaaaaaaaaaaaaa 1199


RESULT   2
AAH14942
ID   AAH14942 standard; cDNA; 1208 BP.
XX
AC   AAH14942;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:12845.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;  99JP-0248036.
PR   27-AUG-1999;  99JP-0300253.
PR   11-JAN-2000;  2000JP-0118776.
PR   02-MAY-2000;  2000JP-0183767.
PR   09-JUN-2000;  2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 12845; 2537pp + CD ROM; English.
XX
```

```
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 1208 BP; 326 A; 295 C; 337 G; 250 T; 0 other;

  Query Match             94.4%;  Score 1131.4;  DB 22;  Length 1208;
  Best Local Similarity   98.2%;  Pred. No. 1.8e-291;
  Matches 1172;  Conservative    0;  Mismatches    1;  Indels   20;  Gaps    2;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
Db      21 gtgacaggagccgaagcagcagcgcaggttgtccccgtttccccctccccttcccttctc 80

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
Db      81 cggttgccttcccgggcccccttacactccacagtcccggtcccgccatgtcccagaaaca 140

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
Db     141 agaagaagagaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaaga 200

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
Db     201 aggtattctgcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctagg 260

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
Db     261 acaaaagcctggaggctccgacttcctcatgaagagactccagaagggcaaaagtactt 320

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
Db     321 tgactcaggagactacaacatggccaaagccaagatgaagaataagcagctgccaagtgc 380

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
Db     381 aggaccagacaagaacctggtgactggtgatcacatccccaccccacaggatctgcccca 440

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480
Db     441 gagaaagtcctcgctcgtcaccagcaagcttgcgggtggccaagttgaatgatgctgccc 500

Qy     481 GGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTC 540
Db     501 ggggctctgccagatcctgaga----------cgcttcccctccctgccccacccgggtc 550

Qy     541 CTGTGCTGGCTCCTGCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTG 590
Db     551 ctgtgctggctcctgccccttcctgcttttgcagccaggggtcaggaggtggctcgggtg 610

Qy     591 TGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGC 650
Db     611 tgggctggagaggcagaagccctttcctgttggtgtcccagcacatggagcccccttgggc 670

Qy     651 TGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGA 710
Db     671 tgagcaccaagaccttgaacctttttttgttttacctttttttccaaataacagttgggaga 730

Qy     711 AATATCAATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGGA 770
Db     731 aatatcaatgaaattctgggggtgggggtggggttgaaagggtggggtgggagatatgga 790

Qy     771 GGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACG 830
Db     791 ggagtatgaattagggcttggagttggtaaaaacattcctgactatccttcttaaccacg 850

Qy     831 TGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTC 890
Db     851 tggctgatgtggggtaggtatgaggggaaggaagtggagtagcctaatgaaaaggggttc 910

Qy     891 TAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATA 950
Db     911 tagttgagctctgtagataaatgccttgtttcagtgtggttggagacctggtgtcagata 970

Qy     951 AAAGAAACTCCATCCGCACAGACAGATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAGTT 1010
Db     971 aaagaaactccatccgcacagacagatgcaaacagctcctctagttctgcagagctagtt 1030

Qy    1011 GAGAACTCAACATTAATCATTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGGGAA 1070
Db    1031 gagaactcaacattaatcattttaaaaagtactgtccttgaaatagatttgctgtgggaa 1090

Qy    1071 GAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAA 1130
Db    1091 gaagggcagtgagtgtgggagaaaggagccgtgagcgtggggaaccccacagagcccaaa 1150

Qy    1131 GGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAA 1183
Db    1151 ggacttttcagtattcgaaataaacaaaacaaaaacccatgaaaaaacccaa 1203

RESULT   3
AAH07746
ID   AAH07746 standard; cDNA; 808 BP.
XX
AC   AAH07746;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:4581.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
```

```
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 1; SEQ ID 4581; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 808 BP; 198 A; 218 C; 224 G; 159 T; 9 other;

  Query Match            47.6%;  Score 570.8;  DB 22;  Length 808;
  Best Local Similarity  94.1%;  Pred. No. 4.3e-142;
  Matches  664;  Conservative    0;  Mismatches   16;  Indels   26;  Gaps    6;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 gtgacaggagccgaagcagcagcgcaggttgtccccgtttccctccccctttcccttctc 80

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 cggttgccttcccgggccccttacactccacagtcccggtcccgccatgtcccagaaaca 140

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 agaagaagagaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaaga 200

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 aggtattctgcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctagg 260

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 acaaaagcctggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactt 320

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 tgactcaggagactacaacatggccaaagccaagatgaagaataagcagctgccaagtgc 380

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 aggaccagacaagaacctggtgactggtgatcacatccccaccccacaggatctgcccca 440

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     441 gagaaagtcctcgctcgtcaccagcaagcttgcgggtggccaagttgaatgatgctgccc 500

Qy     481 GGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCC-ACCCGGGT 539
          ||||||||||||||||||||          ||||||||||||||||| | ||| ||||
Db     501 ggggctctgccagatcctgaga----------cgcttcccctccctgccncaaccggggt 550

Qy     540 CCTGTGCTGGCTCCTGCCCCTTC----------AGCCAGGGGTCAGGAGGTGGCTCGGG 588
          |||||||||||||||||||||||          ||||||||||||||||||||||||||
```

```
Db     551 cctgtgctggctcctgccccttcctgcttttgcaagccaggggtcaggaggtggctcggg 610

Qy     589 TGT--GGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTT 646
           |||  ||   |   ||||| || |||||||||||||||||| ||| ||||||||||||||
Db     611 tgttgggctggagaaggcanaancccttcctgttggtgt-ccancacatggagcccctt 669

Qy     647 GGGCTGAGCACC-AAGACCTTGAACCTTTTTTGTTTTACCTTTTTT 691
           |||||||||||| |||||||||| | ||||||||||  | |||||
Db     670 gggctgagcaccaaagaccttgaaacctttttgtttttacctttttt 715


RESULT   4
AAH12719/c
ID   AAH12719 standard; cDNA; 587 BP.
XX
AC   AAH12719;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (3'-primer) SEQ ID NO:9554.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 3; SEQ ID 9554; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
```

```
XX
SQ   Sequence 587 BP; 143 A; 166 C; 89 G; 175 T; 14 other;


  Query Match          43.0%;  Score 515.6;  DB 22;  Length 587;
  Best Local Similarity  93.0%;  Pred. No. 2e-127;
  Matches  543;  Conservative    0;  Mismatches   38;  Indels    3;  Gaps    1;

Qy    600 GAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCA 659
          | || |  | || || | ||| ||| |||||  | | ||  ||| ||||     |   ||
Db    586 GGGGGAANANCCTTTNCNTGTGGGTNTCCCAAAAAAGGGCCCCCTTTGG---NTGGCACA 530

Qy    660 AGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAAT 719
          |   | || | |||||||||| || ||||||||||||||||||||||||||||||||||||
Db    529 AAGACNTGGAACTTTTTTGTTTAACTTTTTTTCCAAATAACAGTTGGGAGAAATATCAAT 470

Qy    720 GAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGGAGGAGTATGA 779
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    469 GAAATTCTGGGGGTGGGGGTGGGGTTGAAAGGGTGGGGTGGGAGATATGGAGGAGTATGA 410

Qy    780 ATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACGTGGCTGATG 839
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    409 ATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTNTTAACCACGTGGCTGATG 350

Qy    840 TGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTAGTTGAGC 899
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    349 TGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTNTAGTTGAGN 290

Qy    900 TCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATAAAAGAAACT 959
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    289 TCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATAAAAGAAACT 230

Qy    960 CCATCCGCACAGACAGATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAGTTGAGAACTCA 1019
          |||||||||||||||||||||||| |||| ||||| ||||||||||||||||||||||
Db    229 CCATCCGCACAGACAGATGCAAACAGNTCCTNTAGTTNTGCAGAGCTAGTTGAGAACTCA 170

Qy   1020 ACATTAATCATTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGGGAAGAAGGGCAG 1079
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    169 ACATTAATCATTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGGGAAGAAGGGCAG 110

Qy   1080 TGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTT 1139
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    109 TGAGTGTGGGAGAAAGGANCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTT 50

Qy   1140 CAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAA 1183
          |||||||||||||||||||||||||||||||||||||||||||
Db     49 CAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAA 6


RESULT   5
ABK15891
ID   ABK15891 standard; cDNA; 553 BP.
XX
AC   ABK15891;
XX
DT   14-MAR-2002  (first entry)
XX
DE   Human lung tumour protein encoding cDNA #3.
XX
KW   Human; lung tumour protein; lung cancer; immunostimulant; cytostatic; ss;
KW   PCR primer.
XX
OS   Homo sapiens.
XX
PN   WO200202623-A2.
XX
PD   10-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20975.
XX
PR   29-JUN-2000; 2000US-215696P.
PR   22-AUG-2000; 2000US-227143P.
PR   06-SEP-2000; 2000US-230484P.
PR   21-DEC-2000; 2000US-257729P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Mcneill PD,  Wantanabe Y,  Carter D,  Henderson RA,  Kalos MD;
XX
DR   WPI; 2002-106602/14.
XX
PT   Polynucleotides encoding lung tumour polypeptides, useful for treating
PT   lung cancer or stimulating an immune response -
XX
PS   Claim 1; Page 133; 277pp; English.
XX
CC   The invention relates to polynucleotides encoding lung tumour
CC   polypeptides. Compositions comprising the lung tumour polypeptides,
CC   polynucleotides, fusion proteins, antibodies to the polypeptides, T cell
CC   populations, or antigen presenting cells that express the lung tumour
CC   polypeptides, are useful for treating lung cancer or stimulating an
CC   immune response. The polynucleotides and polypeptides are also useful in
CC   the diagnosis and monitoring of lung cancer. Sequences ABK15889-AAS16343
CC   represent cDNA molecules encoding the polypeptides of the invention and
CC   PCR primers used to amplify the cDNA.
XX
SQ   Sequence 553 BP; 151 A; 164 C; 146 G; 92 T; 0 other;


  Query Match          42.0%;  Score 503;  DB 24;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 4.4e-124;
  Matches  503;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     43 gtgacaggagccgaagcagcagcgcaggttgtccccgtttccccteccccttccttctc 102

Qy     61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    103 cggttgccttcccgggcccctacactccacagtcccggtcccgccatgtcccagaaaca 162

Qy    121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    163 agaagaagagaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaaga 222

Qy    181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    223 aggtattctgcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctagg 282

Qy    241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    283 acaaaagcctggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactt 342

Qy    301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    343 tgactcaggagactacaacatggccaaagccaagatgaagaataagcagctgccaagtgc 402

Qy    361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    403 aggaccagacaagaacctggtgactggtgatcacatccccaccccacaggatctgcccca 462

Qy    421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    463 gagaaagtcctcgctcgtcaccagcaagcttgcgggtggccaagttgaatgatgctgccc 522

Qy    481 GGGGCTCTGCCAGATCCTGAGAC 503
          ||||||||||||||||||||||
Db    523 ggggctctgccagatcctgagac 545


RESULT   6
AAZ24516
ID   AAZ24516 standard; cDNA; 669 BP.
XX
AC   AAZ24516;
XX
```

DT 07-DEC-1999 (first entry)
XX
DE Human lung tumor associated polynucleotide.
XX
KW Human; lung tumor; lung cancer; T cell stimulation; ss.
XX
OS Homo sapiens.
XX
PN WO9947674-A2.
XX
PD 23-SEP-1999.
XX
PF 17-MAR-1999; 99WO-US05798.
XX
PR 18-MAR-1998; 98US-0040802.
PR 18-MAR-1998; 98US-0040984.
PR 27-JUL-1998; 98US-0123912.
PR 27-JUL-1998; 98US-0123933.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Wang T;
XX
DR WPI; 1999-571839/48.
XX
PT New isolated lung tumor polynucleotides, used to develop products for
PT the treatment, prevention and monitoring the progression of lung cancer
PT -
XX
PS Claim 1; Page 66; 148pp; English.
XX
CC The invention provides isolated human lung tumor nucleic acids and
CC polypeptides. The polypeptides can be used for the treatment of lung
CC cancer. The polypeptides and polynucleotides can be used to stimulate T
CC cells or antigen presenting cells for use in the treatment of lung
CC cancer. The polypeptides and monoclonal antibodies specific for the
CC polypeptides can also be used to inhibit the development of lung cancer.
CC Agents which bind the polypeptides can be used for detecting lung cancer
CC and for monitoring the progression of lung cancer.
XX
SQ Sequence 669 BP; 158 A; 208 C; 173 G; 119 T; 11 other;

Query Match 41.9%; Score 501.8; DB 20; Length 669;
Best Local Similarity 90.6%; Pred. No. 9.9e-124;
Matches 616; Conservative 0; Mismatches 37; Indels 28; Gaps 7;

```
Qy      10 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 gccgaagcagcagcgcaggttgtccccgtttccccctcccccttcccttctccggttgcct 61

Qy      70 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 tcccgggccccttacactccacagtcccggtcccgccatgtcccagaaacaagaagaaga 121

Qy     130 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 gaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaagaaggtattct 181

Qy     190 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 gcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctaggacaaaagcc 241

Qy     250 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 309
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     242 tggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactttgactcngg 301

Qy     310 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 369
           ||||||||||||||||||||||| |||||||||||||||||||||||||| |||||||
Db     302 agactacaacatggccaaagccaacatgaagaataagcagctgccaagtgcangaccaga 361

Qy     370 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 429
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     362 caagaacctggtgactggtgatcacatccccaccccacaggatctg-cccagagaaagtc 420

Qy     430 CTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTG 489
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     421 ctcgctcgtcaccagcaagcttgcgggtggccaagttgaatgatgctg-ccggggctctg 479

Qy     490 CCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGG 549
           ||| |||            |       | ||||||||||||||||||||||||||||||||
Db     480 ccanatc------------tgagacgcttccctccctgccccacccgggtcctgtgctgg 527

Qy     550 CTCCTGCCC---------CTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAG 600
           |||||||||         | |||||| ||||||||| ||||| | |||   |||||||
Db     528 ctcctgcccttcctgcttttgcagccangggtcaggaagtggcncnggtngtggctggaa 587

Qy     601 AGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAA 660
           ||    || |||||||||||||||||||| | |||||||||||| ||||      |||
Db     588 ag--caaaaccctttcctgttggtgtccca-cccatggagcccctggggc--gagcccan 642

Qy     661 GACCTTGAACCTTTTTTGTTT 681
           || ||||| |||||||  | |
Db     643 gaacttgancctttttgttnt 663
```

RESULT 7
AAC65755
ID AAC65755 standard; cDNA; 669 BP.
XX
AC AAC65755;
XX
DT 21-FEB-2001 (first entry)
XX
DE Human lung cancer-associated cDNA LST-S2-72.
XX
KW Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW vaccine; detection; ss.
XX
OS Homo sapiens.
XX
PN WO200061612-A2.
XX
PD 19-OCT-2000.
XX
PF 03-APR-2000; 2000WO-US08896.
XX
PR 02-APR-1999; 99US-0285479.
PR 17-DEC-1999; 99US-0466396.
PR 30-DEC-1999; 99US-0476496.
PR 10-JAN-2000; 2000US-0480884.
PR 22-FEB-2000; 2000US-0510376.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Fan L;
XX
DR WPI; 2000-628399/60.
XX
PT Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT protein is used for detecting and monitoring progression of lung cancer
PT in a patient -
XX
PS Claim 1a; Page 99; 261pp; English.
XX
CC This invention describes a novel isolated polypeptide (I) which
CC comprising an immunogenic portion of a lung tumor protein or variant (P2)
CC which have cytostatic activity. The polypeptides and polynucleotides are
CC used in compositions and vaccines to inhibit the development of cancer,
CC especially lung cancer, in a patient. Methods described in the invention
CC can be used to monitor the progression of a cancer by carrying out the
CC detection at subsequent time points and comparing the results from the
CC different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
CC are treated with P2, polynucleotides encoding P2 or antigen presenting
CC cells expressing P2 and then administered to the patient to inhibit
CC development of cancer.

```
XX
SQ   Sequence 669 BP; 158 A; 208 C; 173 G; 119 T; 11 other;


  Query Match             41.9%;  Score 501.8;  DB 21;  Length 669;
  Best Local Similarity   90.5%;  Pred. No. 9.9e-124;
  Matches  616;  Conservative    0;  Mismatches   37;  Indels   28;  Gaps    7;

Qy      10 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 gccgaagcagcagcgcaggttgtccccgtttcccctcccccttcccttctccggttgcct 61

Qy      70 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 tcccgggccccttacactccacagtcccggtcccgccatgtcccagaaacaagaagaaga 121

Qy     130 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 gaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaagaaggtattct 181

Qy     190 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 gcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctaggacaaaagcc 241

Qy     250 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db     242 tggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactttgactcngg 301

Qy     310 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 369
           |||||||||||||||||||||| |||||||||||||||||||||||||||| |||||||
Db     302 agactacaacatggccaaagccaacatgaagaataagcagctgccaagtgcangaccaga 361

Qy     370 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 429
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     362 caagaacctggtgactggtgatcacatccccaccccacaggatctg-cccagagaaagtc 420

Qy     430 CTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTG 489
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     421 ctcgctcgtcaccagcaagcttgcgggtggccaagttgaatgatgctg-ccggggctctg 479

Qy     490 CCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGG 549
           ||| |||          |       | ||||||||||||||||||||||||||||||||
Db     480 ccanatc-----------tgagacgcttcctccctgccccacccgggtcctgtgctgg 527

Qy     550 CTCCTGCCC---------CTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAG 600
           |||||||||          | |||||| |||||||| ||||| | |||   |||||||
Db     528 ctcctgcccttcctgcttttgcagccangggtcaggaagtggcncnggtngtggctggaa 587

Qy     601 AGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAA 660
           ||    || |||||||||||||||||| | ||||||||||| ||||      |||
Db     588 ag--caaaaccctttcctgttggtgtccca-cccatggagcccctggggc--gagcccan 642

Qy     661 GACCTTGAACCTTTTTTGTTT 681
           || ||||| ||||||| | |
Db     643 gaacttgancctttttgttnt 663


RESULT   8
AAC01495
ID   AAC01495 standard; cDNA; 503 BP.
XX
AC   AAC01495;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 1493.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG01489.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 1493; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 503 BP; 143 A; 147 C; 133 G; 78 T; 2 other;


  Query Match             38.5%;  Score 461.2;  DB 21;  Length 503;
  Best Local Similarity   99.6%;  Pred. No. 5.8e-113;
  Matches  460;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      31 gtgacaggagccgaagcagcagcgcaggttgtccccgtttcccctcccccttcccttctc 90

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91 cggttgccttcccgggcccctacactccacagtcccggtcccgccatgtcccagaaaca 150

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151 agaagaagagaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaaga 210

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     211 aggtattctgcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctagg 270

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     271 acaaaagcctggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactt 330

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     331 tgactcaggagactacaacatggccaaagccaagatgaagaataagcagctgccaagtgc 390

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
           |||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db     391 aggaccagacaagaacctggtgactggtgatcacatccccamccccacaggatctgcccca 450

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCA 462
           |||||||||||||||||||:||||||||||||||||||||
Db     451 gagaaagtcctcgctcgtcamcagcaagcttgcgggtggcca 492
```

```
RESULT   9
AAV45210
ID   AAV45210 standard; cDNA; 1103 BP.
XX
AC   AAV45210;
XX
DT   19-OCT-1998  (first entry)
XX
DE   Human endosulfine A gene.
XX
KW   ds; endosulfine; central nervous system; pancreatic beta-cell; pain;
KW   insulin; cytoprotective agent; diabetes; epilepsy; depression; ischemia.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             107..460
FT                   /*tag= a
FT                   /product= "Endosulfine A protein"
XX
PN   WO9830692-A2.
XX
PD   16-JUL-1998.
XX
PF   07-JAN-1998;   98WO-US00137.
XX
PR   07-JAN-1997;   97US-0779775.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Anderson KL,  Roch J,  Scott VE,  Sullivan JP;
XX
DR   WPI; 1998-399140/34.
DR   P-PSDB; AAW61551.
XX
PT   Novel human endosulfine A and B - for screening for compounds
PT   modulating endosulfine receptor activity, useful as cytoprotective
PT   agent(s)
XX
PS   Claim 2; Page 57-58; 76pp; English.
XX
CC   The human endosulfine A and B genes encode proteins related to other
CC   members of the endosulfine family which have been shown to bind receptors
CC   from both the central nervous system and pancreatic beta-cells and to
CC   induce insulin secretion from a rodent beta-cell tumour line.  The
CC   protein can be used to screen for compounds that modulate endosulfine
CC   receptor activity, especially as a cytoprotective agent.
CC   Modulators of endosulfine activity may represent a treatment of
CC   diabetes, epilepsy, pain, depression and ischemia.
XX
SQ   Sequence 1103 BP; 313 A; 264 C; 267 G; 259 T; 0 other;

  Query Match             38.0%;  Score 456;  DB 19;  Length 1103;
  Best Local Similarity   100.0%;  Pred. No. 2e-111;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gtgacaggagccgaagcagcagcgcaggttgtccccgtttccccteccccttcccttctc 60

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 cggttgccttcccgggccccttacactccacagtcccggtcccgccatgtcccagaaaca 120

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 agaagaagagaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaaga 180

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     181 aggtattctgcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctagg 240

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 acaaaagcctggaggctccgacttcctcatgaagagactccagaagggcaaaagtactt 300

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 tgactcaggagactacaacatggccaaagccaagatgaagaataagcagctgccaagtgc 360

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 aggaccagacaagaacctggtgactggtgatcacatccccaccccacaggatctgcccca 420

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
          |||||||||||||||||||||||||||||||||||
Db     421 gagaaagtcctcgctcgtcaccagcaagcttgcggg 456


RESULT  10
ABK15892
ID   ABK15892 standard; cDNA; 494 BP.
XX
AC   ABK15892;
XX
DT   14-MAR-2002  (first entry)
XX
DE   Human lung tumour protein encoding cDNA #4.
XX
KW   Human; lung tumour protein; lung cancer; immunostimulant; cytostatic; ss;
KW   PCR primer.
XX
OS   Homo sapiens.
XX
PN   WO200202623-A2.
XX
PD   10-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20975.
XX
PR   29-JUN-2000; 2000US-215696P.
PR   22-AUG-2000; 2000US-227142P.
PR   06-SEP-2000; 2000US-230481P.
PR   21-DEC-2000; 2000US-257729P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Mcneill PD,  Wantanabe Y,  Carter D,  Henderson RA,  Kalos MD;
XX
DR   WPI; 2002-106602/14.
XX
PT   Polynucleotides encoding lung tumour polypeptides, useful for treating
PT   lung cancer or stimulating an immune response -
XX
PS   Claim 1; Page 133; 277pp; English.
XX
CC   The invention relates to polynucleotides encoding lung tumour
CC   polypeptides. Compositions comprising the lung tumour polypeptides,
CC   polynucleotides, fusion proteins, antibodies to the polypeptides, T cell
CC   populations, or antigen presenting cells that express the lung tumour
CC   polypeptides, are useful for treating lung cancer or stimulating an
CC   immune response. The polynucleotides and polypeptides are also useful in
CC   the diagnosis and monitoring of lung cancer. Sequences ABK15889-AAS16343
CC   represent cDNA molecules encoding the polypeptides of the invention and
CC   PCR primers used to amplify the cDNA.
XX
SQ   Sequence 494 BP; 142 A; 146 C; 127 G; 79 T; 0 other;

  Query Match             37.7%;  Score 452;  DB 24;  Length 494;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-110;
  Matches  452;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43 gtgacaggagccgaagcagcagcgcaggttgtccccgtttcccctcccccttcccttctc 102

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103 cggttgccttcccgggcccccttacactccacagtcccggtcccgccatgtcccagaaaca 162

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     163 agaagaagagaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaaga 222

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     223 aggtattctgcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctagg 282

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     283 acaaaagcctggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactt 342

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     343 tgactcaggagactacaacatggccaaagccaagatgaagaataagcagctgccaagtgc 402

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     403 aggaccagacaagaacctggtgactggtgatcacatccccaccccacaggatctgcccca 462

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTG 452
           ||||||||||||||||||||||||||||||||
Db     463 gagaaagtcctcgctcgtcaccagcaagcttg 494
```

RESULT 11
AAC98847
ID AAC98847 standard; cDNA; 863 BP.
XX
AC AAC98847;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:75.
XX
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS Homo sapiens.
XX
PN WO200055320-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05989.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-579444/54.
DR P-PSDB; AAB54082.
XX
PT New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX
PS Claim 1; Page 549-550; 1379pp; English.
XX
CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 863 BP; 236 A; 215 C; 223 G; 186 T; 3 other;

Query Match 35.1%; Score 420.8; DB 21; Length 863;
Best Local Similarity 98.7%; Pred. No. 4.4e-102;
Matches 444; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

```
Qy       8 GAGCCGAAGCAGCAGCGCAGGTTGT-CCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTG 66
           |  |  ||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       1 gnccgnaagcagcagcgcaggttgtaccccgtttcccctcccccttcccttctccggttg 60

Qy      67 CCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGA 126
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ccttcccgggcccccttacactccacagtcccggtcccgccatgtcccagaaacaagaaga 120

Qy     127 AGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTAT 186
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     121 agagaaccctgcggaggagaccggcgaggagaagcaggacacgca-gagaaagaaggtat 179

Qy     187 TCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAA 246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180 tctgcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctaggacaaaa 239

Qy     247 GCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTC 306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 gcctggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactttgactc 299

Qy     307 AGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACC 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 aggagactacaacatggccaaagccaagatgaagaataagcagctgccaagtgcaggacc 359

Qy     367 AGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAA 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 agacaagaacctggtgactggtgatcacatccccaccccacaggatctgccccagagaaa 419

Qy     427 GTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
           |||||||||||||||||||||||||||||
Db     420 gtcctcgctcgtcaccagcaagcttgcggg 449
```

RESULT 12
AAV30285
ID AAV30285 standard; DNA; 454 BP.
XX
AC AAV30285;
XX
DT 28-SEP-1998 (first entry)
XX
DE Human phosphoprotein hPSHP DNA.

XX
KW hPSHP; phosphoprotein; neurotransmitter; signal transduction;
KW human; Parkinson's disease; Huntington's diseases; diagnosis;
KW drug screening; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 17..373
FT /*tag= a
XX
PN WO9818922-A1.
XX
PD 07-MAY-1998.
XX
PF 14-OCT-1997; 97WO-US18489.
XX
PR 28-OCT-1996; 96US-0739484.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Hillman JL;
XX
DR WPI; 1998-272217/24.
DR P-PSDB; AAW60214.
XX
PT Human phosphoprotein polypeptide, hPSHP - useful for treating
PT degenerative neuronal diseases e.g. Parkinson's and Huntington's,
PT diagnosing hPSHP related diseases and drug screening
XX
PS Claim 5; Page 49; 64pp; English.
XX
CC This polynucleotide codes for hPSHP (see AAW60214) a novel human
CC phosphoprotein that is believed to function in the signal
CC transduction pathway of neurotransmitters in brain tissue and in
CC cell development processes in peripheral organs, based on its
CC chemical and structural homology with known phosphoproteins
CC GI 162690 and GI 741603. It is a consensus sequence derived from
CC overlapping and/or extended Incyte clones 258512, 483293, 483619,
CC 485751 and 486742 (from hNT2 teratocarcinoma cell line cDNA library
CC HNT2RAT01), 841590 (PROSNOT05), 1613311 (COLNTUT06), 918496
CC (BRSTNOT04), 1518226 (BLADTUT04) and 1509755 (LUNGNOT14). Incyte
CC clone 25812 was initially identified from a retinoic acid treated
CC HNT2RAT01 library through a computer-generated search for amino acid
CC sequence alignments. The invention also provides genetically
CC engineered expression vectors, and host cells comprising nucleic
CC acid sequences encoding hPSHP used in a claimed method for
CC producing hPSHP, as well as probes and primers, antibodies,
CC agonists and antagonists of hPSHP. hPSHP or its fragments are
CC useful therapeutically to regenerate and enhance survival of nerve
CC cells to halt degenerative processes in brain diseases such as
CC Parkinson's and Huntington's disease. Polynucleotides encoding
CC hPSHP are also useful therapeutically, e.g. to create antisense
CC molecules to block the biological activity of hPSHP to treat
CC diseases associated with hPSHP expression. They may be used
CC diagnostically e.g. to detect and quantify gene expression in
CC biopsied tissues in which hPSHP expression is implicated (e.g.
CC associated with neurological diseases) and to monitor hPSHP levels
CC during therapeutic intervention. The probes can also be used for
CC detection of polynucleotide sequences encoding hPSHP in biological
CC samples by hybridisation assay (the sample may be subjected to PCR
CC prior to analysis) (claimed). The products may also be used in
CC drug screening.
XX
SQ Sequence 454 BP; 120 A; 119 C; 129 G; 85 T; 1 other;

Query Match 28.9%; Score 346.4; DB 19; Length 454;
Best Local Similarity 96.5%; Pred. No. 2.2e-82;
Matches 389; Conservative 0; Mismatches 2; Indels 12; Gaps 3;

```
Qy   162 AGGACACGCAGGAG-AAAGAAGGTATTCTGCCTGAG-AGAGCTGAAGAGGCAAAGCTAAA 219
         ||||||||||||||  |||||||||||||||||||||  |||||||||||||||||||||||
Db    61 aggacacgcaggagcaaagaaggtattctgcctgagaagagctgaagaggcaaagctaaa 120

Qy   220 GGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACT 279
         ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   121 ggccaaatacccaagcctaggacaaaagcctggaggctccgactttctcatgaagagact 180

Qy   280 CCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAA 339
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ccagaaagggcaaaagtactttgactcaggagactacaacatggccaaagccaagatgaa 240

Qy   340 GAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCC 399
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 gaataagcagctgccaagtgcaggaccagacaagaacctggtgactggtgatcacatccc 300

Qy   400 CACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGG 459
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 caccccacaggatctgccccagagaaagtcctcgctcgtcaccagcaagcttgcgggtgg 360

Qy   460 CCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCC 519
         ||||||||||||||||||||||||||||||||||||||||||||          |||||||
Db   361 ccaagttgaatgatgctgcccggggctctgccagatcctgaga----------cgcttcc 410

Qy   520 CCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTC 562
         ||||||||| |||||||||||||||||||||||||||||||||
Db   411 cctccctgcnccacccgggtcctgtgctggctcctgcccccttc 453
```

RESULT 13
AAC06352
ID AAC06352 standard; cDNA; 312 BP.
XX
AC AAC06352;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 10427.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 10427; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

```
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 312 BP; 105 A; 58 C; 77 G; 71 T; 1 other;


  Query Match             25.1%;  Score 300.6;  DB 21;  Length 312;
  Best Local Similarity   99.4%;  Pred. No. 3.1e-70;
  Matches  311;  Conservative    1;  Mismatches    0;  Indels    1;  Gaps    1;

Qy    869 GTAGCCTAATGAAAAGGGGTTCTAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAGTGTG 928
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 gtagcctaatgaaaaggggttctagttgagctctgtagataaatgccttgtttcagtgtg 60

Qy    929 GTTGGAGACCTGGTGTCAGATAAAAGAAACTCCATCCGCACAGACAGATGCAAACAGCTC 988
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 gttggagacctggtgtcagataaaagaaactccatccgcacagacagatgcaaacagctc 120

Qy    989 CTCTAGTTCTGCAGAGCTAGTTGAGAACTCAACATTAATCATTTTAAAAAGTACTGTCCT 1048
          |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ckctagttctgcagagctagttgagaactcaacattaatcattttaaaaagtactgtcct 180

Qy   1049 TGAAATAGATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGT 1108
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 tgaaatagatttgctgtgggaagaagggcagtgagtgtgggagaaaggagccgtgagcgt 240

Qy   1109 GGGGAACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACC 1168
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ggggaa-cccacagagcccaaaggacttttttcagtattcgaaataaacaaaacaaaaacc 299

Qy   1169 CATGAAAAAACCC 1181
          |||||||||||||
Db    300 catgaaaaaaccc 312


RESULT  14
AAQ60964/c
ID   AAQ60964 standard; DNA; 314 BP.
XX
AC   AAQ60964;
XX
DT   16-MAR-1994  (first entry)
XX
DE   Human brain Expressed Sequence Tag EST01042.
XX
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
OS   Homo sapiens.
XX
PN   WO9316178-A.
XX
PD   19-AUG-1993.
XX
PF   12-FEB-1993;   93WO-US01294.
XX
PR   12-FEB-1992;   92US-0837195.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Adams MD,  Moreno RF,  Venter CJ;
XX
DR   WPI; 1993-272882/34.
XX
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
XX
PS   Example 4; Page 418; 500pp; English.
XX
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST01042 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM. See also AAQ59041-Q61440.
XX
SQ   Sequence 314 BP; 72 A; 79 C; 57 G; 106 T; 0 other;


  Query Match             23.8%;  Score 285.8;  DB 14;  Length 314;
  Best Local Similarity   98.7%;  Pred. No. 2.7e-66;
  Matches  309;  Conservative    0;  Mismatches    2;  Indels    2;  Gaps    2;

Qy    865 TGGAGTAGCCTAATG-AAAAGGGGTTCTAGTTGAGCTCTGTAGAT-AAATGCCTTGTTTC 922
          ||||||||||||||| |||||||||||||||||||||||||||| |||||||||| ||
Db    313 TGGAGTAGCCTAATGAAAAAGGGGTTCTAGTTGAGCTCTGTAGATAAAATGCCTTGCATC 254

Qy    923 AGTGTGGTTGGAGACCTGGTGTCAGATAAAAGAAACTCCATCCGCACAGACAGATGCAAA 982
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    253 AGTGTGGTTGGAGACCTGGTGTCAGATAAAAGAAACTCCATCCGCACAGACAGATGCAAA 194

Qy    983 CAGCTCCTCTAGTTCTGCAGAGCTAGTTGAGAACTCAACATTAATCATTTTAAAAAGTAC 1042
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    193 CAGCTCCTCTAGTTCTGCAGAGCTAGTTGAGAACTCAACATTAATCATTTTAAAAAGTAC 134

Qy   1043 TGTCCTTGAAATAGATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGT 1102
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    133 TGTCCTTGAAATAGATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGT 74

Qy   1103 GAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACA 1162
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     73 GAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACA 14

Qy   1163 AAAACCCATGAAA 1175
          |||||||||||||
Db     13 AAAACCCATGAAA 1


RESULT  15
AAQ60963
ID   AAQ60963 standard; DNA; 343 BP.
XX
AC   AAQ60963;
XX
DT   16-MAR-1994  (first entry)
XX
DE   Human brain Expressed Sequence Tag EST01041.
XX
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal;
KW   cAMP-regulated phosphoprotein; signal transduction; ss.
XX
OS   Homo sapiens.
XX
PN   WO9316178-A.
XX
PD   19-AUG-1993.
XX
PF   12-FEB-1993;   93WO-US01294.
XX
PR   12-FEB-1992;   92US-0837195.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Adams MD,  Moreno RF,  Venter CJ;
XX
DR   WPI; 1993-272882/34.
XX
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
```

```
XX
PS   Example 4; Page 418; 500pp; English.
XX
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST01041 was found to have 86.4 per cent identity with a
CC   cAMP-regulated phosphoprotein in the PIR database and has been
CC   classified in the "Signal Transduction" group of ESTs.
CC   See also Q59041-Q61440.
XX
SQ   Sequence 343 BP; 62 A; 104 C; 92 G; 84 T; 1 other;

  Query Match             23.6%;  Score 283;  DB 14;  Length 343;
  Best Local Similarity   92.6%;  Pred. No. 1.6e-65;
  Matches  326;  Conservative    0;  Mismatches    6;  Indels   20;  Gaps    2;

Qy     390 ATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGC 449
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 atcacttccccaccccacaggatctgccccagagaaagtcctcgctcgtcaccagcaagc 60

Qy     450 TTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTT 509
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ttgcgggtggccaagttgaatgatgctgccccggggctctgccagatcctgaga------- 113

Qy     510 TGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTT-------- 561
           |||||||||||||||||||||||||||||||||||| ||||||||||
Db     114 ---cgcttcccctccctgccccacccgggtcctgtgctggntcctgccccttcctgcttt 170

Qy     562 --CAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTG 619
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171 tgcagccaggggtcaggaggtggctcgggtgtgggctggagaggcagaagccctttcctg 230

Qy     620 TTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGT 679
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     231 ttggtgtcccagcacatggagcccctgggctgagcaccaagaccttgaacctttttgt 290

Qy     680 TTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGG 731
           |||||||||||||||||||||||| ||||||||||||||||||   ||||
Db     291 tttacctttttccaaataacagtttggagaaatatcaatgaaatctggggg 342


RESULT  16
AAI93443
ID   AAI93443 standard; cDNA; 2498 BP.
XX
AC   AAI93443;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 13503.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO13512.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 13503; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2498 BP; 735 A; 454 C; 464 G; 845 T; 0 other;

  Query Match             14.1%;  Score 168.6;  DB 22;  Length 2498;
  Best Local Similarity   70.5%;  Pred. No. 1.2e-34;
  Matches  225;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy     138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
           | ||||   ||  || |||| ||||| |      |||  |  |||  | |   || ||||
Db      56 ccgaggcagcctccgcggaggagcagaaggaaatggaagataaagtgactagtccagaga 115

Qy     198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
           ||| ||||| |||||  |||| || | ||| ||    || ||||||||||||||||| |
Db     116 aagcagaagaagcaaaattaaaagcaagatatcctcatctgggacaaaagcctggaggtt 175

Qy     258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
           | || ||| | | |||  |  | ||||||||||||||| || ||||| || || || ||||
Db     176 cagatttcttaaggaaacggttgcagaaagggcaaaaatattttgattctggggattaca 235

Qy     318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
           ||||||| ||||| || |||||||| ||||| || || | |||||  || || ||||
Db     236 acatggctaaagcaaaaatgaagaacaagcaacttcctactgcagctccggataagacgg 295

Qy     378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
           ||| |||||||| ||||| ||||| || || || || || ||  | ||| | || || |
Db     296 aggtcactggtgaccacattcccactccgcaagaccttcctcaacggaagccgtcccttg 355

Qy     438 TCACCAGCAAGCTTGCGGG 456
           |  | |||||||| || ||
Db     356 ttgctagcaagctggctgg 374


RESULT  17
AAK90275/c
ID   AAK90275 standard; DNA; 2544 BP.
XX
AC   AAK90275;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 3851.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
```

XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01324.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX

PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
PS Disclosure; SEQ ID NO 3851; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.
XX
SQ Sequence 2544 BP; 698 A; 653 C; 506 G; 687 T; 0 other;

Query Match 13.8%; Score 165.4; DB 22; Length 2544;
Best Local Similarity 96.6%; Pred. No. 8.9e-34;
Matches 169; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2531 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 2472

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2471 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 2412

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAG 175
          ||||||||||||||||||||||||||||||||||||||||||  | |  | ||
Db    2411 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGTGAAGTGGAAG 2357
```

RESULT 18
AAI57655/c
ID AAI57655 standard; DNA; 2544 BP.
XX
AC AAI57655;
XX
DT 19-OCT-2001 (first entry)
XX
DE Human colorectal cancer antigen coding sequence SEQ ID NO: 192.
XX
KW Human; colorectal cancer; colorectal cancer antigen; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155350-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01350.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.

```
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-457727/49.
XX
PT   Isolated polypeptide for treating, preventing and/ or prognosing
PT   disorders related to the colon and rectum including colorectal cancers
PT   and also for testing and detection e.g. diagnosis -
XX
PS   Disclosure; SEQ ID NO: 192; 522pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of colorectal cancer antigens. These are shown in
CC   AAI57547-AAI57619 and AAM38569-AAM38641. These can be used in the
CC   diagnosis, prevention and treatment of cancer of the colon and/or rectum.
CC   The present sequence is a colorectal cancer antigen genomic sequence.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2544 BP; 698 A; 653 C; 506 G; 687 T; 0 other;
```

```
  Query Match             13.8%;  Score 165.4;  DB 22;  Length 2544;
  Best Local Similarity   96.6%;  Pred. No. 8.9e-34;
  Matches  169;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2531 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 2472

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2471 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 2412

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAG 175
          ||||||||||||||||||||||||||||||||||||||||  | |  | ||
Db    2411 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGTGAAGTGGAAG 2357


RESULT  19
AAV38840
ID   AAV38840 standard; DNA; 3316 BP.
XX
AC   AAV38840;
XX
DT   27-OCT-1998  (first entry)
XX
DE   Homo sapiens C4-2 gene.
XX
KW   C4-2; tumour suppressor gene; cancer; treatment; glioblastoma;
KW   cell proliferative disorders; brain tumour; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..291
FT                   /*tag= a
FT                   /product= C4-2 tumour suppressor
XX
PN   WO9820021-A1.
XX
PD   14-MAY-1998.
XX
PF   07-NOV-1997;   97WO-US20408.
XX
PR   08-NOV-1996;   96US-0744905.
XX
PA   (NWBI-) NORTHWEST BIOTHERAPEUTICS LLC.
XX
PI   Boynton AL,  Murphy GP,  Sehgal A;
XX
DR   WPI; 1998-286852/25.
DR   P-PSDB; AAW62655.
XX
PT   Tumour suppressor, C4-2 - useful for treatment of proliferative
PT   diseases e.g. cancer
XX
PS   Example; Fig 8; 112pp; English.
XX
CC   The sequence is that of the C4-2 tumour suppressor gene.
CC   The gene, its antisense sequences and its encoded protein
CC   can be used in the treatment of cell proliferative diseases,
CC   e.g. cancer, in particular glioblastomas and other malignant
CC   brain tumours.
XX
SQ   Sequence 3316 BP; 1010 A; 565 C; 581 G; 1160 T; 0 other;
```

```
  Query Match             13.0%;  Score 155.4;  DB 19;  Length 3316;
  Best Local Similarity   71.6%;  Pred. No. 4.6e-31;
  Matches  204;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy     172 GGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCC 231
```

```
Db      3 ggaagataaagtgactagtccagagaaagcagaagaagcaaaattaaaagcaagatatcc 62

Qy    232 AAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCA 291
Db     63 tcatctgggacaaaagcctggaggttcaaatttcttaaggaaacggttgcagaaagggca 122

Qy    292 AAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCT 351
Db    123 aaaatattttgattctggggattacaacatggctaaagcaaaaatgaagaacaagcaact 182

Qy    352 GCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGA 411
Db    183 tcctactgcagctccggataagacggaggtcactggtgaccacattcccactccgcaaga 242

Qy    412 TCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
Db    243 ccttcctcaactgaagccgtcccttgttgctagcaagctggctgg 287


RESULT  20
AAC00778
ID   AAC00778 standard; cDNA; 341 BP.
XX
AC   AAC00778;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 776.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG00772.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 776; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 341 BP; 102 A; 76 C; 99 G; 61 T; 3 other;


  Query Match             10.2%;  Score 121.8;  DB 21;  Length 341;
  Best Local Similarity   70.7%;  Pred. No. 1.5e-22;
  Matches  162;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy    138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
Db    112 ccgaggcagcctccgcggaggagcagaaggaaatggaagataaagtgactagtccagaga 171

Qy    198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
Db    172 aagcagaagaagcaaaattaaaagcaagatatcctcatctgggacaaaagcctggaggtt 231

Qy    258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
Db    232 cagatttcttaaggaaacggttgcagaaagggcaaaaatattttgattctggggattaca 291

Qy    318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACC 366
Db    292 acatggctaaagcaaaaatgaagaacaagcaacttcctactgcagctcc 340


RESULT  21
AAC00776
ID   AAC00776 standard; cDNA; 350 BP.
XX
AC   AAC00776;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 774.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG00770.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 774; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
```

```
SQ   Sequence 350 BP; 124 A; 63 C; 106 G; 57 T; 0 other;


  Query Match            10.2%;  Score 121.8;  DB 21;  Length 350;
  Best Local Similarity   70.7%;  Pred. No. 1.5e-22;
  Matches  162;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
           | ||||   ||  || |||| ||||| |      ||| | ||| | |   || ||||
Db     121 ccgaggcagcctccgcggaggagcagaaggaaatggaagataaagtgactagtccagaga 180

Qy     198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
           ||| ||||| |||||  |||| || | ||| ||    || ||||||||||||||||| |
Db     181 aagcagaagaagcaaaattaaaagcaagatatcctcatctgggacaaaagcctggaggtt 240

Qy     258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
           | || ||| | | |||  |  | |||||||||||||| || ||||| || || || ||||
Db     241 cagatttcttaaggaaacggttgcagaaagggcaaaaatattttgattctggggattaca 300

Qy     318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACC 366
           ||||||| ||||| || |||||||| |||| || || | |||||  ||
Db     301 acatggctaaagcaaaaatgaagaacaagcaacttcctactgcagctcc 349


RESULT  22
AAC00775
ID   AAC00775 standard; cDNA; 295 BP.
XX
AC   AAC00775;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 773.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG00769.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 773; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 295 BP; 111 A; 48 C; 85 G; 51 T; 0 other;


  Query Match             9.5%;  Score 114;  DB 21;  Length 295;
  Best Local Similarity   68.8%;  Pred. No. 1.7e-20;
  Matches  172;  Conservative    0;  Mismatches   75;  Indels    3;  Gaps    1;

Qy     120 AAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGG---ACACGCAGGAGA 176
           | ||| ||  |||   |  ||| || | ||    |||||||| ||   |      |||
Db      45 aggaaaaaaggaagtgtttggacgacagcgagcgggagaagcggggtcaagaaatggaag 104

Qy     177 AAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCC 236
           |  |||  | |   || |||| ||| ||||| |||||  |||| || | ||| ||    |
Db     105 ataaagtgactagtccagagaaagcagaagaagcaaaattaaaagcaagatatcctcatc 164

Qy     237 TAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGT 296
           | |||||||||||||||| || || ||| | | |||  |  | ||||||||||||| |
Db     165 tgggacaaaagcctggaggttcagatttcttaaggaaacggttgcagaaagggcaaaaat 224

Qy     297 ACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAA 356
           | ||||| || || || |||||||||| ||||| || |||||||| ||||| || || |
Db     225 attttgattctggggattacaacatggctaaagcaaaaatgaagaacaagcaacttccta 284

Qy     357 GTGCAGGACC 366
           |||||  ||
Db     285 ctgcagctcc 294


RESULT  23
AAC00777
ID   AAC00777 standard; cDNA; 407 BP.
XX
AC   AAC00777;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 775.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG00771.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 775; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
```

```
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 407 BP; 140 A; 81 C; 116 G; 70 T; 0 other;


  Query Match             9.3%;  Score 111.8;  DB 21;  Length 407;
  Best Local Similarity   73.3%;  Pred. No. 7.5e-20;
  Matches  143;  Conservative    0;  Mismatches   52;  Indexes    0;  Gaps    0;

Qy    172 GGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCC 231
          ||| | ||| | | || |||| ||| ||||| ||||| |||| || | ||| ||
Db    212 ggaagataaagtgactagtccagagaaagcagaagaagcaaaattaaaagcaagatatcc 271

Qy    232 AAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCA 291
          || |||||||||||||||| || || ||| | | |||  |  | ||||||||||
Db    272 tcatctgggacaaaagcctggaggttcagatttcttaaggaaacggttgcagaaagggca 331

Qy    292 AAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCT 351
          ||| || ||||| || || || |||||||||| ||||| || |||||||| ||||| ||
Db    332 aaaatattttgattctggggattacaacatggctaaagcaaaaatgaagaacaagcaact 391

Qy    352 GCCAAGTGCAGGACC 366
           || | |||||  ||
Db    392 tcctactgcagctcc 406


RESULT  24
AAC15418
ID   AAC15418 standard; cDNA; 91 BP.
XX
AC   AAC15418;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 19493.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 19493; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 91 BP; 13 A; 33 C; 20 G; 14 T; 11 other;


  Query Match             5.7%;  Score 67.8;  DB 21;  Length 91;
  Best Local Similarity   90.1%;  Pred. No. 2e-08;
  Matches   64;  Conservative    7;  Mismatches    0;  Indexes    0;  Gaps    0;

Qy      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          |:|||||||||||||||||||||||||::|||||||||||||::||||||||||:|||
Db     21 gkgacaggagccgaagcagcagcgcagghwgtccccgtttccccwmcccctttccctwctc 80

Qy     61 CGGTTGCCTTC 71
          |||:|||||||
Db     81 cggwtgccttc 91


RESULT  25
AAX27245
ID   AAX27245 standard; DNA; 1835 BP.
XX
AC   AAX27245;
XX
DT   28-MAY-1999  (first entry)
XX
DE   Human CLAR1 coding sequence.
XX
KW   CLAR1; human; late stage-specific marker; prostate cancer progression;
KW   cancer susceptibility; diagnosis; therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9909215-A1.
XX
PD   25-FEB-1999.
XX
PF   13-AUG-1998;   98WO-US16768.
XX
PR   13-AUG-1997;   97US-0055285.
XX
PA   (FOXC-) FOX CHASE CANCER CENT.
XX
PI   Rondinelli R,  Tricoli JV;
XX
DR   WPI; 1999-181061/15.
DR   P-PSDB; AAY00922, AAY00923.
XX
PT   Newly isolated CLAR1 nucleic acid molecule (including complements
PT   and allelic variants) encoding a polypeptide - useful as a late
PT   stage-specific marker for the prognosis and diagnosis of prostate
PT   cancer progression
XX
PS   Claim 1; Page 43-44; 70pp; English.
XX
CC   This sequence encodes the human CLAR1 of the invention. CLAR1 was
CC   isolated from a human adult heart cDNA library and is a late
CC   stage-specific marker for prostate cancer progression. The CLAR1 nucleic
CC   acid molecules (including oligonucleotides as probes/primers) and
CC   antibodies against CLAR1 form kits for detecting expression of the CLAR1
CC   gene associated with a susceptibility to cancer. The nucleic acids are
CC   also useful for identifying and characterising related genes from other
CC   species. They are useful for identifying genes encoding proteins that
CC   interact with CLAR1, which facilitates further study of the role of these
CC   cellular signalling mechanisms in cancer progression. The nucleic acids
CC   are also useful for controlling CLAR1 protein production. The CLAR1
```

```
CC   polypeptides are useful in screening for molecules (especially
CC   antibodies) that affect or modulate its activity or function. Such
CC   molecules may be therapeutically useful. The isolated CLAR1 nucleic acid
CC   molecule enables the generation of a transgenic animal, which can be
CC   used as a model for prostate cancer to allow screening and identification
CC   of compounds for the treatment of prostate cancer. Such an animal
CC   expresses the CAR1 protein at a higher level than that of a normal
CC   animal.
XX
SQ   Sequence 1835 BP; 384 A; 607 C; 514 G; 330 T; 0 other;


  Query Match             4.0%;  Score 47.6;  DB 20;  Length 1835;
  Best Local Similarity   61.1%;  Pred. No. 0.02;
  Matches   77;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy   1074 GGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGA 1133
           ||| || ||||| |||     ||   ||  ||   || |   | | | |||   ||||
Db   1678 ggggagagagtgaggggttgggggatcggcagttgggaggggcgctctgagattaaagag 1737

Qy   1134 CTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAAA 1193
           |||  |  |     ||| ||| |||| |||||   |  ||||||  ||||||||||||
Db   1738 ttttacctttggggtaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1797

Qy   1194 AAAAAA 1199
           ||||||
Db   1798 aaaaaa 1803


RESULT  26
AAZ52551
ID   AAZ52551 standard; cDNA; 551 BP.
XX
AC   AAZ52551;
XX
DT   29-FEB-2000  (first entry)
XX
DE   Human secreted protein clone ya80_1 nucleotide sequence SEQ ID NO:153.
XX
KW   Human; secreted protein; immunostimulatory; haemostatic; cytokine;
KW   proliferative; differentiative; chemotactic; chemokinetic; vaccine;
KW   thrombolytic; antiinflammatory; cytostatic; immunosuppressive;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9958642-A2.
XX
PD   18-NOV-1999.
XX
PF   14-MAY-1999;   99WO-US10843.
XX
PR   14-MAY-1998;   98US-0085472.
PR   17-AUG-1998;   98US-0096824.
PR   11-SEP-1998;   98US-0099843.
PR   11-SEP-1998;   98US-0099950.
PR   15-SEP-1998;   98US-0100424.
PR   29-SEP-1998;   98US-0102329.
PR   09-OCT-1998;   98US-0103615.
PR   11-DEC-1998;   98US-0111799.
PR   14-DEC-1998;   98US-0112159.
PR   31-DEC-1998;   98US-0114415.
PR   10-FEB-1999;   99US-0248059.
PR   06-APR-1999;   99US-0287150.
PR   13-MAY-1999;   99US-0311021.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Wong GG,  Clark HF,  Fechtel K,  Agostino MJ;
XX
DR   WPI; 2000-053095/04.
DR   P-PSDB; AAY73466.
XX
PT   Novel polynucleotides and proteins having biological activities which
PT   make them suitable for treating, preventing or ameliorating medical
PT   conditions in humans or animals -
XX
PS   Claim 162; Page 678; 730pp; English.
XX
CC   The present invention describes human secreted proteins encoded by
CC   polynucleotides obtained from adult testes, foetal brain, adult brain,
CC   brain (foetal and adult), foetal kidney, adult spleen, and adult thymus
CC   cDNA libraries. The polynucleotides and proteins are predicted to have
CC   biological activities which would make them suitable for treating,
CC   preventing or ameliorating medical conditions in humans and animals.
CC   Suggested activities include nutritional activity, cytokine and cell
CC   proliferation/differentiation activity, immune stimulating (e.g. as
CC   vaccines) or suppressing activity, haematopoiesis regulating activity,
CC   tissue growth activity, activin/inhibin activity, chemotactic/
CC   chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC   ligand activity, anti-inflammatory activity, cadherin/tumour invasion
CC   suppressor activity, and tumour inhibition activity. The polynucleotides
CC   are also stated to be useful for gene therapy. Therapeutic compositions
CC   are also presently valuable for veterinary applications. AA252475 to
CC   AAZ52581 encode human secreted proteins, and AAY73390 to AAY73500
CC   represent human secreted proteins, given in the present invention.
XX
SQ   Sequence 551 BP; 173 A; 120 C; 155 G; 103 T; 0 other;


  Query Match             4.0%;  Score 47.4;  DB 21;  Length 551;
  Best Local Similarity   48.7%;  Pred. No. 0.013;
  Matches  129;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy    933 GAGACCTGGTGTCAGATAAAAGAAACTCCATCCGCACAGACAGATGCAAACAGCTCCTCT 992
          || ||  |     |||  |   ||||||  |  |  || |    |  |   |||||
Db    287 gacacaagacaagagagcaccaaaactctctaggagcacaggaagtccttcagctttcag 346

Qy    993 AGTTCTGCAGAGCTAGTTGAGAACTCAACATTAATCATTTTAAAAAGTACTGTCCTTGAA 1052
            || | |||| |     ||  || || | |||  | ||  | |||  | |    |||
Db    347 gaatcagaagaggtgtcacagctctgaagaataaggacttccacaagcccaggagatgag 406

Qy   1053 ATAGATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGG 1112
           |  |  | |  ||. ||| |   || | || |||  |  || |      | |   | | |
Db    407 aaggggtggaagttggaggggcgggactgggtggagtggacattccaggagctgggtgag 466

Qy   1113 AACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATG 1172
               |||  |    | | |  |    |||  |    | |    ||| |||||   |
Db    467 gtaaaaacacggaggcaggaaaatgcggagtgctttggaaaccataaaaaaaaaaaaaaa 526

Qy   1173 AAAAAACCCAAAAAAAAAAAAAAAA 1197
           ||||||   ||||||||||||||||
Db    527 aaaaaaaaaaaaaaaaaaaaaaaaa 551


RESULT  27
AAS01015
ID   AAS01015 standard; cDNA; 1133 BP.
XX
AC   AAS01015;
XX
DT   18-MAY-2001  (first entry)
XX
DE   Maize disease resistance enhancing protein ZmLSD1(2)-1 cDNA.
XX
KW   Maize; plant disease resistance; crop; soybean; sunflower; sorghum;
KW   canola; wheat; alfalfa; cotton; rice; barley; millet; plant cell death;
KW   herbicide resistance; ZmLSD1(2)-1; lesions stimulating disease; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..161
FT                   /*tag=  a
FT   CDS             162..698
FT                   /*tag=  b
```

```
FT                   /product= "ZmLSD1(2)-1 protein"
FT   3'UTR           699..1064
FT                   /*tag=  c
FT   polyA_site      1065
FT                   /*tag=  d
XX
PN   WO200118061-A2.
XX
PD   15-MAR-2001.
XX
PF   06-SEP-2000; 2000WO-US24403.
XX
PR   09-SEP-1999;   99US-0152988.
XX
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Simmons CR;
XX
DR   WPI; 2001-226742/23.
DR   P-PSDB; AAU00445.
XX
PT   Novel isolated maize disease resistance polynucleotide useful for
PT   increasing resistance in a plant to disease, controlling cell death,
PT   and conferring resistance to herbicides -
XX
PS   Claim 1; Page 74-75; 90pp; English.
XX
CC   The present sequence encoding for maize ZmLSD1(2)-1 protein is 1 of 7
CC   novel disease resistance proteins (AAU00443-AAU00449). The ZmLSD1(2)-1 is
CC   a Zea mays gene homologue to Arabidopsis LSD1 gene (lesions stimulating
CC   disease) which contains a leucine zipper. These novel maize disease
CC   resistance polynucleotides and polypeptides are useful for enhancing
CC   disease resistance in crops and transgenic plants including maize,
CC   soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice,
CC   barley or millet. The level of disease resistance protein is increased
CC   by transforming a plant cell with a recombinant expression cassette
CC   which comprises the disease resistance polynucleotide operably linked to
CC   a promoter, or by culturing the plant cell under plant growing conditions
CC   to produce a regenerated plant, or by inducing expression of the
CC   polynucleotide to modulate the disease resistance protein in a plant.
CC   The polynucleotides encoding the disease resistance proteins are useful
CC   for increasing resistance in a plant to disease, controlling cell death,
CC   and conferring resistance to herbicides. They are useful as probes or
CC   amplification primers in the detection, quantification, or isolation
CC   of gene transcripts. They can be used for recombinant expression of
CC   their encoded polypeptides, as immunogens in the preparation and/or
CC   screening of antibodies, and in sense or antisense suppression of the
CC   polynucleotide in a host cell, tissue or plant.
XX
SQ   Sequence 1133 BP; 313 A; 259 C; 271 G; 290 T; 0 other;
```

```
  Query Match             3.9%;  Score 47;  DB 22;  Length 1133;
  Best Local Similarity   60.6%;  Pred. No. 0.023;
  Matches   77;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy   1073 AGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGG 1132
          | | | ||| | |  |  |    | ||| ||    | | |||   || | |   |||| |
Db    994 aagcccgtgggcgctgttgctgcgtgccttgttattagtgaataacaaacattgcaaatg 1053

Qy   1133 ACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAA 1192
           || ||  | ||    ||| ||| |||| |||||   |  ||||||   ||||||||||
Db   1054 gctcttgaaataaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1113

Qy   1193 AAAAAAA 1199
          |||||||
Db   1114 aaaaaaa 1120
```

```
RESULT  28
AAX26196
ID   AAX26196 standard; DNA; 1613 BP.
XX
AC   AAX26196;
XX
DT   25-MAY-1999  (first entry)
XX
DE   DNA sequence of glycosyltransferase gene.
XX
KW   Glycosyltransferase gene; beta1,3-galactosyltransferase;
KW   GD1/GM1/GA1 synthase; ss.
XX
OS   Unidentified.
XX
PN   JP11056373-A.
XX
PD   02-MAR-1999.
XX
PF   27-AUG-1997;   97JP-0246176.
XX
PR   27-AUG-1997;   97JP-0246176.
XX
PA   (FURU/) FURUKAWA K.
XX
DR   WPI; 1999-222386/19.
DR   P-PSDB; AAW95171.
XX
PT   A glycosyltransferase gene - having beta1,3-galactosyltransferase
PT   activity
XX
PS   Claim 4; Page 9-10; 14pp; Japanese.
XX
CC   The present sequence represents a glycosyltransferase gene encoding a
CC   protein exhibiting beta1,3-galactosyltransferase activity. The invention
CC   shows an identity of GD1/GM1/GA1 synthase gene.
XX
SQ   Sequence 1613 BP; 385 A; 464 C; 450 G; 314 T; 0 other;
```

```
  Query Match             3.9%;  Score 46.2;  DB 20;  Length 1613;
  Best Local Similarity   61.0%;  Pred. No. 0.044;
  Matches   75;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy   1077 CAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTT 1136
          ||  ||    ||  | | | |||   | | ||||||||  |   |||||||||   ||
Db   1473 caaagaccagggagggacgaagcacagggtgtggggaagactcaagagcccaataaacac 1532

Qy   1137 TTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAAAAAA 1196
           |          ||| ||| |||| |||||   |  ||||||   ||||||||||||||
Db   1533 gtctctctacaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1592

Qy   1197 AAA 1199
          |||
Db   1593 aaa 1595
```

```
RESULT  29
AAI84623
ID   AAI84623 standard; cDNA; 405 BP.
XX
AC   AAI84623;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 4683.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
```

```
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO04692.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 4683; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 405 BP; 192 A; 72 C; 63 G; 77 T; 1 other;


  Query Match             3.8%;  Score 45.6;  DB 22;  Length 405;
  Best Local Similarity   71.4%;  Pred. No. 0.034;
  Matches   60;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy   1116 CCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAA 1175
            ||| ||  |||   | ||||||||||  |    ||| ||| ||||| ||||   |  |||
Db     57 ccccccatggcctccatgacttttttcaaaaaaaaaaaaaaaaaaaaccaaaaaaaaaaaaa 116

Qy   1176 AAACCCAAAAAAAAAAAAAAAAAA 1199
            |||   ||||||||||||||||||
Db    117 aaaaaaaaaaaaaaaaaaaaaaa 140


RESULT  30
AAI84652
ID   AAI84652 standard; cDNA; 453 BP.
XX
AC   AAI84652;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 4712.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO04721.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 4712; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 453 BP; 210 A; 79 C; 68 G; 92 T; 4 other;


  Query Match             3.8%;  Score 45.6;  DB 22;  Length 453;
  Best Local Similarity   71.4%;  Pred. No. 0.035;
  Matches   60;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy   1116 CCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAA 1175
            ||| ||  |||   | ||||||||||  |    ||| ||| ||||| ||||   |  |||
Db     57 ccccccatggcctccatgacttttttcaaaaaaaaaaaaaaaaaaaaccaaaaaaaaaaaaa 116

Qy   1176 AAACCCAAAAAAAAAAAAAAAAAA 1199
            |||   ||||||||||||||||||
Db    117 aaaaaaaaaaaaaaaaaaaaaaa 140


RESULT  31
AAC78104
ID   AAC78104 standard; cDNA; 840 BP.
XX
AC   AAC78104;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human cancer associated gene sequence SEQ ID NO:498.
XX
KW   Human; cancer associated gene; cancer antigen; detection; cancer;
KW   diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW   antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW   antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW   dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW   vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW   immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW   allergic reaction; graft versus host disease; organ rejection;
KW   haemostatic; thrombolytic; cardiovascular disorder; infection;
KW   neurological disease; drug screening; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055350-A1.
XX
PD   21-SEP-2000.
```

XX
PF 08-MAR-2000; 2000WO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587533/55.
DR P-PSDB; AAB43895.
XX
PT Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX
PS Claim 1; Page 1029; 2352pp; English.
XX
CC AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerary; immunomodulator;
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 840 BP; 281 A; 169 C; 180 G; 205 T; 5 other;

Query Match 3.8%; Score 45.6; DB 21; Length 840;
Best Local Similarity 56.8%; Pred. No. 0.047;
Matches 84; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy   1052 AATAGATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGG 1111
          |||||  |   |||  |||     | || |     | ||||| | ||  | ||| |
Db    690 aatagtctgtatgtttcaagcccttctgtaaaatatgaagaaaagtgctcttagcattct 749

Qy   1112 GAACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCAT 1171
          |      || |  |    |    | | |  ||| || ||| ||| |||| |||||   |
Db    750 gtgtaaaactgtactgttaaatatatgtgtgtaatcaaaaaaaaaaaaaaaaaaaaaaa 809

Qy   1172 GAAAAAACCCAAAAAAAAAAAAAAAAA 1199
          ||||||   ||||||||||||||||||
Db    810 aaaaaaaaaaaaaaaaaaaaaaaaaaa 837
```

RESULT 32
AAF32706
ID AAF32706 standard; cDNA; 2968 BP.
XX
AC AAF32706;
XX
DT 23-MAR-2001 (first entry)
XX
DE Human secreted protein gene 8 SEQ ID NO:18.
XX
KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;
KW angiogenesis; nervous system disorder; Alzheimer's disease; skin aging;
KW ocular disorder; corneal infection; wound healing; food additive;
KW preservative; ss.
XX
OS Homo sapiens.
XX
PN WO200077255-A1.
XX
PD 21-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US14926.
XX
PR 11-JUN-1999; 99US-0138628.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis GA;
XX
DR WPI; 2001-025337/03.
DR P-PSDB; AAB64429.
XX
PT Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
XX
PS Claim 1; Page 477; 593pp; English.
XX
CC The polynucleotide sequences given in AAF32699 to AAF32747 encode the
CC human secreted proteins given in AAB64422 to AAB64470. AAB64471 to
CC AAB64548 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Examples of activities include:
CC antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
CC cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
CC neuroprotective; antibacterial; virucide; fungicide; and
CC ophthalmological. The polynucleotides and polypeptides can be used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. AAF32690 to AAF32698 and
CC AAB64421 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 2968 BP; 679 A; 816 C; 797 G; 671 T; 5 other;

Query Match 3.8%; Score 45.6; DB 22; Length 2968;
Best Local Similarity 57.9%; Pred. No. 0.083;
Matches 81; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy   1060 TGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCA 1119
          ||||||  | | || |   |||    |     |  ||| || |  |  | |  |||   |
Db   2813 tgctgtcagcataatgattgtgttccttccctatgggatccatctgttttgtaaacaata 2872

Qy   1120 CAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAAC 1179
           || | |  | |||  | |   |  |    ||| ||| |||| |||||   |  ||||||
Db   2873 aagcgtctgagggagtgtaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2932

Qy   1180 CCAAAAAAAAAAAAAAAAAA 1199
          ||||||||||||||||||||
```

```
Db   2933 aaaaaaaaaaaaaaaaaaa 2952


RESULT  33
AAI95102/c
ID   AAI95102 standard; cDNA; 794 BP.
XX
AC   AAI95102;
XX
DT   13-NOV-2001  (first entry)
XX
DE   Human neuroblastoma expressed polynucleotide SEQ ID NO 1177.
XX
KW   Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS   Homo sapiens.
XX
PN   WO200166719-A1.
XX
PD   13-SEP-2001.
XX
PF   02-MAR-2001; 2001WO-JP01629.
XX
PR   07-MAR-2000; 2000JP-0159195.
XX
PA   (CHIB-) CHIBA PREFECTURE.
PA   (HISM ) HISAMITSU PHARM CO LTD.
XX
PI   Nakagawara A;
XX
DR   WPI; 2001-565584/63.
XX
PT   Nucleic acids originating in gene expressed in human neuroblastoma,
PT   useful as probe or primer in diagnosing prognosis of human
PT   neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT   for anti-cancer agents  -
XX
PS   Claim 1; Page 894; 2979pp; Japanese.
XX
CC   The invention relates to novel genes (AAI93926-AAI97963) expressed in
CC   human neuroblastoma. The nucleic acids are applicable as a probe or
CC   primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC   susceptibility indicators or tumour markers for anti-cancer agents. The
CC   gene information for diagnosing prognosis is related to factors similar
CC   to that for N-myc and TrkA genes.
XX
SQ   Sequence 794 BP; 134 A; 140 C; 186 G; 162 T; 172 other;


  Query Match             3.7%;  Score 44.8;  DB 22;  Length 794;
  Best Local Similarity   60.4%;  Pred. No. 0.074;
  Matches   61;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy   1099 CCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAA 1158
          || |  | |   ||   ||| | |  |||       |||||   |     ||| ||| ||
Db    174 CCTTNGGGGGNNGGNNTCCCCCGGNCCCCCNNCCNTTTTTTNNNTTNAAAAAAAAAAAAA 115

Qy   1159 AACAAAAACCCATGAAAAAACCCAAAAAAAAAAAAAAAAAA 1199
          || |||||    |  ||||||   ||||||||||||||||||
Db    114 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 74


RESULT  34
AAL26422
ID   AAL26422 standard; cDNA; 2115 BP.
XX
AC   AAL26422;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 18879.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer  -
XX
PS   Claim 1; Page 3468-3469; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 2115 BP; 564 A; 430 C; 464 G; 645 T; 12 other;


  Query Match             3.7%;  Score 44.8;  DB 22;  Length 2115;
  Best Local Similarity   76.4%;  Pred. No. 0.12;
  Matches   55;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy   1128 AAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAA 1187
          ||| |  | ||||||| | ||||| ||| |||| |||||   |  ||||||  ||||||
Db   2027 aaaaggatatttcagtgtgcgaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2086

Qy   1188 AAAAAAAAAAAA 1199
          ||||||||||||
Db   2087 aaaaaaaaaaaa 2098


RESULT  35
AAI88751
ID   AAI88751 standard; cDNA; 400 BP.
XX
AC   AAI88751;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 8811.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
```

```
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO08820.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 8811; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 400 BP; 189 A; 47 C; 95 G; 69 T; 0 other;

  Query Match             3.7%;  Score 44.6;  DB 22;  Length 400;
  Best Local Similarity   71.1%;  Pred. No. 0.062;
  Matches   59;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy   1117 CCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAA 1176
            || ||  |||   | ||||||||||  |    ||| ||| |||| |||||   |  ||||
Db     58 ccccatggcctgcatgacttttttcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 117

Qy   1177 AACCCAAAAAAAAAAAAAAAAAA 1199
            ||   ||||||||||||||||||
Db    118 aaaaaaaaaaaaaaaaaaaaaa 140


RESULT  36
AAI91888
ID   AAI91888 standard; cDNA; 422 BP.
XX
AC   AAI91888;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 11948.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO11957.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 11948; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 422 BP; 182 A; 91 C; 89 G; 60 T; 0 other;

  Query Match             3.7%;  Score 44.6;  DB 22;  Length 422;
  Best Local Similarity   71.1%;  Pred. No. 0.063;
  Matches   59;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy   1117 CCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAA 1176
            || ||  |||   | ||||||||||  |    ||| ||| |||| |||||   |  ||||
Db     92 ccccatggcctccatgacttttttcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 151

Qy   1177 AACCCAAAAAAAAAAAAAAAAAA 1199
            ||   ||||||||||||||||||
Db    152 aaaaaaaaaaaaaaaaaaaaaa 174


RESULT  37
AAI93088/c
ID   AAI93088 standard; cDNA; 438 BP.
XX
AC   AAI93088;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 13148.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
```

```
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO13157.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 13148; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 438 BP; 94 A; 88 C; 98 G; 158 T; 0 other;


  Query Match             3.7%;  Score 44.6;  DB 22;  Length 438;
  Best Local Similarity   58.8%;  Pred. No. 0.064;
  Matches   77;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy   1069 AAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCA 1128
           || |||||  | | ||     | ||| | |  ||     | |||  ||||  | | |||
Db    181 AAAAAGGGGGGGGGGTAAAAAAAAAAAGGGGGGTTTTTTTAGGGGCCCCCCAAAACCCCT 122

Qy   1129 AAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAA 1188
               |||||          ||| ||| |||| |||||    |  ||||||   |||||||
Db    121 TTTTTTTTTTTTTTCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 62

Qy   1189 AAAAAAAAAAA 1199
           ||||||||||
Db     61 AAAAAAAAAAA 51


RESULT  38
AAL19946/c
ID   AAL19946 standard; cDNA; 457 BP.
XX
AC   AAL19946;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 12403.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US00798.
XX
PR   14-JAN-2000; 2000US-0176077.
PR   14-MAR-2000; 2000US-0189167.
PR   24-MAR-2000; 2000US-0192099.
PR   29-MAR-2000; 2000US-0193480.
PR   15-MAY-2000; 2000US-0205230.
PR   09-JUN-2000; 2000US-0211315.
PR   25-JUL-2000; 2000US-0220534.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer  -
XX
PS   Claim 1; Page 2191-2192; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity.
XX
SQ   Sequence 457 BP; 118 A; 84 C; 96 G; 159 T; 0 other;


  Query Match             3.7%;  Score 44.6;  DB 22;  Length 457;
  Best Local Similarity   71.1%;  Pred. No. 0.065;
  Matches   59;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy   1117 CCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAA 1176
           || |  | || |||  | ||||| | | ||  ||| ||| |||| |||||   |  ||||
Db    136 CCCCCCACCCAAAAAAATTTTTTAAATTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 77

Qy   1177 AACCCAAAAAAAAAAAAAAAAAA 1199
           ||   ||||||||||||||||||
Db     76 AAAAAAAAAAAAAAAAAAAAAA 54


RESULT  39
AAI84684
ID   AAI84684 standard; cDNA; 449 BP.
XX
AC   AAI84684;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 4744.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO04753.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
```

PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 4744; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 449 BP; 172 A; 56 C; 110 G; 109 T; 2 other;

Query Match 3.7%; Score 44.4; DB 22; Length 449;
Best Local Similarity 69.8%; Pred. No. 0.073;
Matches 60; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 1114 ACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGA 1173
Db 22 agcccccatggcctccatgacttttttcaaaaaaaaaaaaaaaaagaaaaaaaaaaaaaaaa 81

Qy 1174 AAAAACCCAAAAAAAAAAAAAAAAAA 1199
Db 82 aaaaaaaaaaaaaaaaaaaaaaaaaa 107

RESULT 40
AAZ52560
ID AAZ52560 standard; cDNA; 2262 BP.
XX
AC AAZ52560;
XX
DT 29-FEB-2000 (first entry)
XX
DE Human secreted protein clone ye7_1 nucleotide sequence SEQ ID NO:171.
XX
KW Human; secreted protein; immunostimulatory; haemostatic; cytokine;
KW proliferative; differentiative; chemotactic; chemokinetic; vaccine;
KW thrombolytic; antiinflammatory; cytostatic; immunosuppressive;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO9958642-A2.
XX
PD 18-NOV-1999.
XX
PF 14-MAY-1999; 99WO-US10843.
XX
PR 14-MAY-1998; 98US-0085472.
PR 17-AUG-1998; 98US-0096824.
PR 11-SEP-1998; 98US-0099843.
PR 11-SEP-1998; 98US-0099950.
PR 15-SEP-1998; 98US-0100424.
PR 29-SEP-1998; 98US-0102329.
PR 09-OCT-1998; 98US-0103615.
PR 11-DEC-1998; 98US-0111799.
PR 14-DEC-1998; 98US-0112159.
PR 31-DEC-1998; 98US-0114415.
PR 10-FEB-1999; 99US-0248059.
PR 06-APR-1999; 99US-0287150.
PR 13-MAY-1999; 99US-0311021.
XX

PA (GEMY) GENETICS INST INC.
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ;
XX
DR WPI; 2000-053095/04.
DR P-PSDB; AAY73475.
XX
PT Novel polynucleotides and proteins having biological activities which
PT make them suitable for treating, preventing or ameliorating medical
PT conditions in humans or animals -
XX
PS Claim 180; Page 686-687; 730pp; English.
XX
CC The present invention describes human secreted proteins encoded by
CC polynucleotides obtained from adult testes, foetal brain, adult brain,
CC brain (foetal and adult), foetal kidney, adult spleen, and adult thymus
CC cDNA libraries. The polynucleotides and proteins are predicted to have
CC biological activities which would make them suitable for treating,
CC preventing or ameliorating medical conditions in humans and animals.
CC Suggested activities include nutritional activity, cytokine and cell
CC proliferation/differentiation activity, immune stimulating (e.g. as
CC vaccines) or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, cadherin/tumour invasion
CC suppressor activity, and tumour inhibition activity. The polynucleotides
CC are also stated to be useful for gene therapy. Therapeutic compositions
CC are also presently valuable for veterinary applications. AAZ52475 to
CC AAZ52581 encode human secreted proteins, and AAY73390 to AAY73500
CC represent human secreted proteins, given in the present invention.
XX
SQ Sequence 2262 BP; 580 A; 620 C; 623 G; 439 T; 0 other;

Query Match 3.7%; Score 44.4; DB 21; Length 2262;
Best Local Similarity 69.8%; Pred. No. 0.15;
Matches 60; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 1114 ACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGA 1173
Db 2074 agcccaaaatgctcaataaagctattcagtgatgaaaaaaaaaaaaaaaaaaaaaaaaaa 2133

Qy 1174 AAAAACCCAAAAAAAAAAAAAAAAAA 1199
Db 2134 aaaaaaaaaaaaaaaaaaaaaaaaaa 2159

Search completed: September 12, 2002, 16:14:37
Job time: 11851 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2002, 16:08:14 ; Search time 104.12 Seconds
(without alignments)
2828.607 Million cell updates/sec

Title: US-09-824-178-2
Perfect score: 1199
Sequence: 1 GTGACAGGAGCCGAAGCAGC.....CCAAAAAAAAAAAAAAAAAA 1199

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	501.8	41.9	669	4	US-09-040-984-16	Sequence 16, Appl
	2	501.8	41.9	669	4	US-09-123-912-16	Sequence 16, Appl
	3	346.4	28.9	454	2	US-08-739-484-2	Sequence 2, Appli
c	4	64.4	5.4	7218	1	US-08-232-463-14	Sequence 14, Appl
	5	42.8	3.6	1129	4	US-09-227-357-40	Sequence 40, Appl
	6	42.6	3.6	2158	1	US-07-602-608-1	Sequence 1, Appli
	7	42.6	3.6	2158	1	US-08-261-578-1	Sequence 1, Appli
	8	42.2	3.5	1558	1	US-08-467-607-2	Sequence 2, Appli
	9	42.2	3.5	1558	2	US-08-469-362-2	Sequence 2, Appli
	10	42.2	3.5	1558	2	US-08-850-392-2	Sequence 2, Appli
	11	42	3.5	991	3	US-08-924-747-25	Sequence 25, Appl
	12	42	3.5	991	4	US-09-247-373B-25	Sequence 25, Appl
	13	42	3.5	991	4	US-09-296-715-25	Sequence 25, Appl
	14	42	3.5	1315	4	US-09-164-193-1	Sequence 1, Appli
	15	41.2	3.4	2184	4	US-08-955-918C-1	Sequence 1, Appli
	16	41.2	3.4	2205	3	US-08-888-077A-41	Sequence 41, Appl
	17	40.6	3.4	1046	1	US-08-361-467B-4	Sequence 4, Appli
	18	40.6	3.4	1046	1	US-08-484-332C-4	Sequence 4, Appli
	19	40.4	3.4	216	1	US-08-686-878A-34	Sequence 34, Appl
	20	40.4	3.4	216	4	US-09-175-928-34	Sequence 34, Appl
c	21	40.2	3.4	612	4	US-09-385-982-280	Sequence 280, App
	22	40.2	3.4	1721	5	PCT-US96-00994-3	Sequence 4, Appli
	23	40	3.3	2301	4	US-09-232-191-8	Sequence 8, Appli
	24	40	3.3	2301	4	US-09-232-200-8	Sequence 8, Appli
	25	40	3.3	2301	4	US-09-232-197-8	Sequence 8, Appli
	26	40	3.3	2301	4	US-09-232-201-8	Sequence 8, Appli
	27	40	3.3	2550	6	5258287-23	Patent No. 5258287
	28	40	3.3	2671	6	5168051-9	Patent No. 5168051
	29	40	3.3	2710	4	US-09-232-200-44	Sequence 44, Appl
	30	40	3.3	2710	4	US-09-232-200-70	Sequence 70, Appl
	31	40	3.3	2710	4	US-09-232-197-44	Sequence 44, Appl
	32	40	3.3	2710	4	US-09-232-197-70	Sequence 70, Appl
	33	40	3.3	2710	4	US-09-232-201-44	Sequence 44, Appl
	34	40	3.3	2710	4	US-09-232-201-70	Sequence 70, Appl
	35	39.8	3.3	2538	3	US-08-899-437-1	Sequence 1, Appli
	36	39.8	3.3	2538	4	US-09-126-121-1	Sequence 1, Appli
c	37	39.6	3.3	629	4	US-09-385-982-389	Sequence 389, App
	38	39.6	3.3	3410	4	US-09-020-956-110	Sequence 110, App
	39	39.6	3.3	3410	4	US-09-030-607-110	Sequence 110, App
	40	39.6	3.3	3410	4	US-09-439-313-110	Sequence 110, App
	41	39.4	3.3	140	1	US-08-628-417-5	Sequence 5, Appli
	42	39.4	3.3	240	1	US-08-628-417-6	Sequence 6, Appli
	43	39.4	3.3	1485	4	US-09-372-422A-39	Sequence 39, Appl
	44	39.4	3.3	1798	4	US-09-797-906-1	Sequence 1, Appli
	45	39.4	3.3	4895	3	US-09-053-866-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-040-984-16
; Sequence 16, Application US/09040984
; Patent No. 6210883
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Wang, TongTong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS
; TITLE OF INVENTION: OF LUNG CANCER
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,984
; FILING DATE: 18-MAR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.456
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-282-6031
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-040-984-16

bad date

Query Match 41.9%; Score 501.8; DB 4; Length 669;
Best Local Similarity 90.5%; Pred. No. 1.5e-135;
Matches 616; Conservative 0; Mismatches 37; Indels 28; Gaps 7;

Qy 10 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 69
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 61

```
Qy      70 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 121

Qy     130 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 181

Qy     190 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 241

Qy     250 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     242 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCNGG 301

Qy     310 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 369
           |||||||||||||||||||| |||||||||||||||||||||||||| |||||||||||
Db     302 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCNGACCAGA 361

Qy     370 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 429
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     362 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTG-CCCAGAGAAAGTC 420

Qy     430 CTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTG 489
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     421 CTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTG-CCGGGGCTCTG 479

Qy     490 CCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGG 549
           ||| |||            |       | ||||||||||||||||||||||||||||||
Db     480 CCANATC------------TGAGACGCTTCCCTCCCTGCCCCACCCGGGTCCTGTGCTGG 527

Qy     550 CTCCTGCCC---------CTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAG 600
           |||||||||          | |||||| |||||||| ||||| | |||   |||||||
Db     528 CTCCTGCCCTTCCTGCTTTTGCAGCCANGGGTCAGGAAGTGGCNCNGGTNGTGGCTGGAA 587

Qy     601 AGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAA 660
           ||    || |||||||||||||||||||| | |||||||||||| ||||      |||
Db     588 AG--CAAAACCCTTTCCTGTTGGTGTCCCA-CCCATGGAGCCCCTGGGGC--GAGCCCAN 642

Qy     661 GACCTTGAACCTTTTTTGTTT 681
           || ||||| |||||||  | |
Db     643 GAACTTGANCCTTTTTGTTNT 663
```

```
RESULT   2
US-09-123-912-16
; Sequence 16, Application US/09123912A
; Patent No. 6312695
; GENERAL INFORMATION:
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Wang, Tongtong
;  TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
;  FILE REFERENCE:  210121.455C1
;  CURRENT APPLICATION NUMBER: US/09/123,912A
;  CURRENT FILING DATE:  1998-07-27
;  PRIOR APPLICATION NUMBER:  09/040,802
;  PRIOR FILING DATE:  1998-03-18
;  NUMBER OF SEQ ID NOS:  114
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 16
;   LENGTH: 669
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (299)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (354)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (483)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (555)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (571)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (573)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (577)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (642)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (651)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (662)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (667)
;   OTHER INFORMATION: Where n is a, c, g or t
US-09-123-912-16
```

```
  Query Match           41.9%;  Score 501.8;  DB 4;  Length 669;
  Best Local Similarity  90.5%;  Pred. No. 1.5e-135;
  Matches  616;  Conservative    0;  Mismatches   37;  Indels   28;  Gaps    7;

Qy      10 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 gccgaagcagcagcgcaggttgtccccgtttccctccccttccttctccggttgcct 61

Qy      70 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 tcccgggccccttacactccacagtcccggtcccgccatgtcccagaaacaagaagaaga 121

Qy     130 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 gaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaagaaggtattct 181

Qy     190 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 gcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctaggacaaaagcc 241

Qy     250 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     242 tggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactttgactcngg 301

Qy     310 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 369
           |||||||||||||||||||| |||||||||||||||||||||||||| |||||||||||
Db     302 agactacaacatggccaaagccaacatgaagaataagcagctgccaagtgcangaccaga 361

Qy     370 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 429
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     362 caagaacctggtgactggtgatcacatccccaccccacaggatctg-cccagagaaagtc 420

Qy     430 CTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTG 489
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     421 ctcgctcgtcaccagcaagcttgcgggtggccaagttgaatgatgctg-ccggggctctg 479

Qy     490 CCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGG 549
           ||| |||            |       | ||||||||||||||||||||||||||||||
Db     480 ccanatc------------tgagacgcttccctccctgccccacccgggtcctgtgctgg 527

Qy     550 CTCCTGCCC---------CTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAG 600
           |||||||||          | |||||| |||||||| ||||| | |||   |||||||
Db     528 ctcctgcccttcctgcttttgcagccanggGTCAGGAAGTGGCNCNGGTNGTGGCTGGAA 587
```

```
Qy     601 AGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAA 660
           ||    || |||||||||||||||||||| | |||||||||||| ||||      |||
Db     588 ag--caaaaccctttcctgttggtgtccca-cccatggagcccctggggc--gagcccan 642

Qy     661 GACCTTGAACCTTTTTTGTTT 681
           || ||||| |||||||  | |
Db     643 gaacttgancctttttgttnt 663


RESULT   3
US-08-739-484-2
; Sequence 2, Application US/08739484
; Patent No. 5917028
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    TITLE OF INVENTION:  A NOVEL HUMAN
;    TITLE OF INVENTION:  PHOSPHOPROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/739,484
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0140 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  454 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:
;      CLONE:  Concensus
US-08-739-484-2


  Query Match             28.9%;  Score 346.4;  DB 2;  Length 454;
  Best Local Similarity   96.5%;  Pred. No. 1e-90;
  Matches  389;  Conservative    0;  Mismatches    2;  Indels   12;  Gaps    3;

Qy     162 AGGACACGCAGGAG-AAAGAAGGTATTCTGCCTGAG-AGAGCTGAAGAGGCAAAGCTAAA 219
           ||||||||||||||| |||||||||||||||||||| |||||||||||||||||||||||
Db      61 AGGACACGCAGGAGCAAAGAAGGTATTCTGCCTGAGAAGAGCTGAAGAGGCAAAGCTAAA 120

Qy     220 GGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACT 279
           ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     121 GGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTTCTCATGAAGAGACT 180

Qy     280 CCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAA 339
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAA 240

Qy     340 GAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCC 399
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     241 GAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCC 300

Qy     400 CACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGG 459
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGG 360

Qy     460 CCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCC 519
           |||||||||||||||||||||||||||||||||||||||||||          |||||||
Db     361 CCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGA----------CGCTTCC 410

Qy     520 CCTCCCTGCCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTC 562
           ||||||||| ||||||||||||||||||||||||||||||||||
Db     411 CCTCCCTGCNCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTC 453


RESULT   4
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             5.4%;  Score 64.4;  DB 1;  Length 7218;
  Best Local Similarity   5.5%;  Pred. No. 1.7e-08;
  Matches   17;  Conservative  186;  Mismatches  107;  Indels    0;  Gaps    0;

Qy     114 AGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGG 173
           ::::: ::::::::::::::    : ::::::::  :: ::::::::: :::: : : :::
Db    1347 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1288
```

```
Qy     174 AGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAA 233
           :::::::::: :    :   :::::::  ::::::: ::::  :::::  ::: :   ::
Db    1287 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1228

Qy     234 GCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAA 293
           :   :::: :::::   :::::    ::       : ::::::::    :::::::: :::
Db    1227 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1168

Qy     294 AGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGC 353
           :: :    ::   :::::::  : :: : ::  ::::  :::: :::::: ::: ::  :
Db    1167 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1108

Qy     354 CAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATC 413
            ::: : ::::  ::: ::::::   :: ::  :: ::  : ||| | | | | | |  || |
Db    1107 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCTCCCTCGACC 1048

Qy     414 TGCCCCAGAG 423
           |||  |  ||
Db    1047 TGCAGCCAAG 1038
```

```
RESULT   5
US-09-227-357-40
; Sequence 40, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
;  APPLICANT: Fischer et al.
;  TITLE OF INVENTION: 123 Human Secreted Proteins
;  FILE REFERENCE: PZ010P1
;  CURRENT APPLICATION NUMBER: US/09/227,357
;  CURRENT FILING DATE: 1999-01-08
;  EARLIER APPLICATION NUMBER: PCT/US98/13684
;  EARLIER FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: 60/051,926
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,793
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,925
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,929
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,803
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,732
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,931
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,932
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,916
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,930
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,918
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,920
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,733
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,795
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,919
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,928
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/055,722
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,723
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,948
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,949
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,953
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,950
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,947
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,964
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/056,360
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,684
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,984
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,954
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/058,785
;  EARLIER FILING DATE: 1997-09-12
;  EARLIER APPLICATION NUMBER: 60/058,664
;  EARLIER FILING DATE: 1997-09-12
;  EARLIER APPLICATION NUMBER: 60/058,660
;  EARLIER FILING DATE: 1997-09-12
;  EARLIER APPLICATION NUMBER: 60/058,661
;  EARLIER FILING DATE: 1997-09-12
;  NUMBER OF SEQ ID NOS: 672
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 40
;   LENGTH: 1129
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1053)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-227-357-40
```

```
  Query Match             3.6%;  Score 42.8;  DB 4;  Length 1129;
  Best Local Similarity   74.6%;  Pred. No. 0.01;
  Matches   53;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy    1129 AAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAA 1188
           ||  | |||||||| | |||  || ||| |||| |||||   |  ||||||   ||||||| 
Db    1033 aataaatttttcatttttccnaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1092

Qy    1189 AAAAAAAAAAA 1199
           |||||||||||
Db    1093 aaaaaaaaaaa 1103
```

```
RESULT   6
US-07-602-608-1
; Sequence 1, Application US/07602608
; Patent No. 5382524
;  GENERAL INFORMATION:
;    APPLICANT:  Desnick, Robert J.
;    APPLICANT:  Bishop, David F.
;    APPLICANT:  Ioannou, Yiannis A.
;    APPLICANT:  Wang, Anne M.
;    TITLE OF INVENTION:  CLONING AND EXPRESSION OF BIOLOGICALLY
;    TITLE OF INVENTION:  ACTIVE ALPHA-N-ACETYLGALACTOSAMINIDASE
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/602,608
;       FILING DATE:  24-OCT-1990
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Coruzzi, Laura A.
;       REGISTRATION NUMBER:  30,742
;       REFERENCE/DOCKET NUMBER:  6923-008
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-8864/9741
;       TELEX:  66141 PENNIE
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2158 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  345..1580
US-07-602-608-1


  Query Match             3.6%;  Score 42.6;  DB 1;  Length 2158;
  Best Local Similarity   67.4%;  Pred. No. 0.017;
  Matches   60;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy    1111 GGAACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCA 1170
           ||  |||    |   ||||   || | |||  || || ||| ||| |||| |||||   |
Db    2056 GGCTCCCTCAGGTCACCAATAAACCTGTTCTTTAATCAAAAAAAAAAAAAAAAAAAAAAA 2115

Qy    1171 TGAAAAAACCCAAAAAAAAAAAAAAAAA 1199
             ||||||   ||||||||||||||||||
Db    2116 AAAAAAAAAAAAAAAAAAAAAAAAAAAA 2144


RESULT   7
US-08-261-578-1
; Sequence 1, Application US/08261578
; Patent No. 5491075
;  GENERAL INFORMATION:
;    APPLICANT:  Desnick, Robert J.
;    APPLICANT:  Bishop, David F.
;    APPLICANT:  Ioannou, Yiannis A.
;    APPLICANT:  Wang, Anne M.
;    TITLE OF INVENTION:  CLONING AND EXPRESSION OF BIOLOGICALLY
;    TITLE OF INVENTION:  ACTIVE ALPHA-N-ACETYLGALACTOSAMINIDASE
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/261,578
;      FILING DATE:  17-JUN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/602,608
;      FILING DATE:  24-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  6923-008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2158 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  345..1580
US-08-261-578-1


  Query Match             3.6%;  Score 42.6;  DB 1;  Length 2158;
  Best Local Similarity   67.4%;  Pred. No. 0.017;
  Matches   60;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy    1111 GGAACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCA 1170
           ||  |||    |   ||||   || | |||  || || ||| ||| |||| |||||   |
Db    2056 GGCTCCCTCAGGTCACCAATAAACCTGTTCTTTAATCAAAAAAAAAAAAAAAAAAAAAAA 2115

Qy    1171 TGAAAAAACCCAAAAAAAAAAAAAAAAA 1199
             ||||||   ||||||||||||||||||
Db    2116 AAAAAAAAAAAAAAAAAAAAAAAAAAAA 2144


RESULT   8
US-08-467-607-2
; Sequence 2, Application US/08467607
; Patent No. 5783434
;  GENERAL INFORMATION:
;    APPLICANT:  TUNG, JAY S.
;    APPLICANT:  SINHA, SUKANTO
;    APPLICANT:  McCONLOGUE, LISA
;    APPLICANT:  TATSUNO, GWEN
;    APPLICANT:  ANDERSON, JOHN
;    APPLICANT:  CHRYSLER, SUSANNA
;    TITLE OF INVENTION:  NOVEL CATHEPSIN AND METHODS AND
;    TITLE OF INVENTION:  COMPOSITIONS FOR INHIBITION THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ATHENA NEUROSCIENCES
;      STREET:  800 F. Gateway Blvd.
;      CITY:  South San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/467,607
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DUVALL, JEAN M.
;      REGISTRATION NUMBER:  32,731
;      REFERENCE/DOCKET NUMBER:  002010-007
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 877-0900
;      TELEFAX:  (415) 877-8370
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1558 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  103..1011
US-08-467-607-2


  Query Match             3.5%;  Score 42.2;  DB 1;  Length 1558;
  Best Local Similarity   51.7%;  Pred. No. 0.018;
  Matches   74;  Conservative    7;  Mismatches   62;  Indels    0;  Gaps    0;

Qy    1057 ATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACC 1116
           ||||    |  |:|  : ||  | |:| | |:| ||  | |      :           ||
Db    1384 ATTTCACYTTAGRATCMAGGGTGGGRGRGRGRGGGAGGGAATTGTCARTTCCCCMTCCCC 1443

Qy    1117 CCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAA 1176
           |||  |     :|| :| | |  |  |   ||||| ||| |||| ||||| |  |||| 
Db    1444 CCANTGNTGRAWAAARAATCTGCCCCTYCCCGAAAAAAAAAAAAAAAAAAAAAAAAAAA 1503

Qy    1177 AACCCAAAAAAAAAAAAAAAAA 1199
           ||  ||||||||||||||||||
Db    1504 AAAAGAAAAAAAAAAAAAAAAA 1526


RESULT   9
US-08-469-362-2
; Sequence 2, Application US/08469362
; Patent No. 5849711
;  GENERAL INFORMATION:
;    APPLICANT:  TUNG, JAY S.
;    APPLICANT:  SINHA, SUKANTO
;    APPLICANT:  McCONLOGUE, LISA
;    APPLICANT:  SEMKO, CHRISTOPHER M.F.
;    TITLE OF INVENTION:  NOVEL CATHEPSIN AND METHODS AND
;    TITLE OF INVENTION:  COMPOSITIONS FOR INHIBITION THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ATHENA NEUROSCIENCES
;      STREET:  800 F. Gateway Blvd.
;      CITY:  South San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,362
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DUVALL, JEAN M.
;      REGISTRATION NUMBER:  32,731
;      REFERENCE/DOCKET NUMBER:  002010-005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 877-0900
;      TELEFAX:  (415) 877-8370
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1558 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  103..1011
US-08-469-362-2


  Query Match             3.5%;  Score 42.2;  DB 2;  Length 1558;
  Best Local Similarity   51.7%;  Pred. No. 0.018;
  Matches   74;  Conservative    7;  Mismatches   62;  Indels    0;  Gaps    0;

Qy    1057 ATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACC 1116
           ||||    |  |:|  : ||  | |:| | |:| ||  | |      :           ||
Db    1384 ATTTCACYTTAGRATCMAGGGTGGGRGRGRGRGGGAGGGAATTGTCARTTCCCCMTCCCC 1443

Qy    1117 CCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAA 1176
           |||  |     :|| :| | |  |  |   ||||| ||| |||| ||||| |  |||| 
Db    1444 CCANTGNTGRAWAAARAATCTGCCCCTYCCCGAAAAAAAAAAAAAAAAAAAAAAAAAAA 1503

Qy    1177 AACCCAAAAAAAAAAAAAAAAA 1199
           ||  ||||||||||||||||||
Db    1504 AAAAGAAAAAAAAAAAAAAAAA 1526


RESULT  10
US-08-850-392-2
; Sequence 2, Application US/08850392
; Patent No. 5858982
;  GENERAL INFORMATION:
;    APPLICANT:  TUNG, JAY S.
;    APPLICANT:  SINHA, SUKANTO
;    APPLICANT:  McCONLOGUE, LISA
;    APPLICANT:  SEMKO, CHRISTOPHER M.F.
;    TITLE OF INVENTION:  NOVEL CATHEPSIN AND METHODS AND
;    TITLE OF INVENTION:  COMPOSITIONS FOR INHIBITION THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ATHENA NEUROSCIENCES
;      STREET:  800 F. Gateway Blvd.
;      CITY:  South San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/850,392
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/469,362
;      FILING DATE:  06-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DUVALL, JEAN M.
;      REGISTRATION NUMBER:  32,731
;      REFERENCE/DOCKET NUMBER:  002010-005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 877-0900
;      TELEFAX:  (415) 877-8370
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1558 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  103..1011
US-08-850-392-2


  Query Match             3.5%;  Score 42.2;  DB 2;  Length 1558;
  Best Local Similarity   51.7%;  Pred. No. 0.018;
  Matches   74;  Conservative    7;  Mismatches   62;  Indels    0;  Gaps    0;

Qy    1057 ATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACC 1116
           ||||    |  |:|  : ||  | |:| | |:| ||  | |      :           ||
Db    1384 ATTTCACYTTAGRATCMAGGGTGGGRGRGRGRGGGAGGGAATTGTCARTTCCCCMTCCCC 1443

Qy    1117 CCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAA 1176
           |||  |     :|| :| | |  |  |   ||||| ||| |||| ||||| |  |||| 
Db    1444 CCANTGNTGRAWAAARAATCTGCCCCTYCCCGAAAAAAAAAAAAAAAAAAAAAAAAAAA 1503

Qy    1177 AACCCAAAAAAAAAAAAAAAAA 1199
           ||  ||||||||||||||||||
Db    1504 AAAAGAAAAAAAAAAAAAAAAA 1526


RESULT  10
US-08-850-392-2
; Sequence 2, Application US/08850392
; Patent No. 5858982
;  GENERAL INFORMATION:
;    APPLICANT:  TUNG, JAY S.
;    APPLICANT:  SINHA, SUKANTO
;    APPLICANT:  McCONLOGUE, LISA
;    APPLICANT:  SEMKO, CHRISTOPHER M.F.
;    TITLE OF INVENTION:  NOVEL CATHEPSIN AND METHODS AND
;    TITLE OF INVENTION:  COMPOSITIONS FOR INHIBITION THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ATHENA NEUROSCIENCES
;      STREET:  800 F. Gateway Blvd.
;      CITY:  South San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/850,392
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/469,362
;      FILING DATE:  06-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DUVALL, JEAN M.
;      REGISTRATION NUMBER:  32,731
;      REFERENCE/DOCKET NUMBER:  002010-005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 877-0900
;      TELEFAX:  (415) 877-8370
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1558 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  103..1011
US-08-850-392-2


  Query Match             3.5%;  Score 42.2;  DB 2;  Length 1558;
  Best Local Similarity   51.7%;  Pred. No. 0.018;
  Matches   74;  Conservative    7;  Mismatches   62;  Indels    0;  Gaps    0;

Qy    1057 ATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACC 1116
           ||||    |  |:|  : ||  | |:| | |:| ||  | |      :           ||
Db    1384 ATTTCACYTTAGRATCMAGGGTGGGRGRGRGRGGGAGGGAATTGTCARTTCCCCMTCCCC 1443
```

```
Qy   1117 CCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAA 1176
          ||| |     :|| :| | |  |  |   ||||| ||| |||| |||||   |  ||||
Db   1444 CCANTGNTGRAWAAARAATCTGCCCCTYCCCGAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1503

Qy   1177 AACCCAAAAAAAAAAAAAAAAA 1199
          ||  ||||||||||||||||||
Db   1504 AAAAGAAAAAAAAAAAAAAAAA 1526


RESULT  11
US-08-924-747-25
; Sequence 25, Application US/08924747
; Patent No. 6063570
;  GENERAL INFORMATION:
;    APPLICANT:  MCGONIGLE, BRIAN
;    APPLICANT:  O'KEEFE, DANIEL
;    TITLE OF INVENTION:  SOYBEAN GLUTATHIONE-S-TRANSFERASE
;    TITLE OF INVENTION:  ENZYMES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E.I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WORD FOR WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD VERSION 7.0A
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/924,747
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CL-1108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  991 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      TISSUE TYPE:  SOYBEAN
;    IMMEDIATE SOURCE:
;      CLONE:  SSM.PK0067.G5
US-08-924-747-25


  Query Match             3.5%;  Score 42;  DB 3;  Length 991;
  Best Local Similarity   73.0%;  Pred. No. 0.016;
  Matches   54;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy  1126 CCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAA 1185
         | |||  |  | ||||||  ||  ||| ||| |||| |||||   |  ||||||  ||||
Db    906 CTAAAAAAAGTGTTCAGTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 965

Qy  1186 AAAAAAAAAAAAAA 1199
         ||||||||||||||
Db    966 AAAAAAAAAAAAAA 979


RESULT  12
US-09-247-373B-25
; Sequence 25, Application US/09247373B
; Patent No. 6168954
; GENERAL INFORMATION:
;  APPLICANT:  MCGONIGLE, BRIAN
;  APPLICANT:  O'KEEFE, DANIEL
;  TITLE OF INVENTION:  SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
;  FILE REFERENCE:  CL-1108-A
;  CURRENT APPLICATION NUMBER: US/09/247,373B
;  CURRENT FILING DATE:  1999-02-10
;  PRIOR APPLICATION NUMBER:  08/924,747
;  PRIOR FILING DATE:  1997-09-05
;  NUMBER OF SEQ ID NOS:  56
;  SOFTWARE:  Microsoft Office 97
; SEQ ID NO 25
;   LENGTH: 991
;   TYPE: DNA
;   ORGANISM: SOYBEAN
US-09-247-373B-25


  Query Match             3.5%;  Score 42;  DB 4;  Length 991;
  Best Local Similarity   73.0%;  Pred. No. 0.016;
  Matches   54;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy  1126 CCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAA 1185
         | |||  |  | ||||||  ||  ||| ||| |||| |||||   |  ||||||  ||||
Db    906 ctaaaaaagtgttcagttttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 965

Qy  1186 AAAAAAAAAAAAAA 1199
         ||||||||||||||
Db    966 aaaaaaaaaaaaaa 979


RESULT  13
US-09-296-715-25
; Sequence 25, Application US/09296715
; Patent No. 6171839
;  GENERAL INFORMATION:
;    APPLICANT:  MCGONIGLE, BRIAN
;    APPLICANT:  O'KEEFE, DANIEL
;    TITLE OF INVENTION:  SOYBEAN GLUTATHIONE-S-TRANSFERASE
;    TITLE OF INVENTION:  ENZYMES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E.I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WORD FOR WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD VERSION 7.0A
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/296,715
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CL-1108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  991 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
```

```
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      TISSUE TYPE:  SOYBEAN
;    IMMEDIATE SOURCE:
;      CLONE:  SSM.PK0067.G5
US-09-296-715-25


  Query Match             3.5%;  Score 42;  DB 4;  Length 991;
  Best Local Similarity   73.0%;  Pred. No. 0.016;
  Matches   54;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy   1126 CCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAA 1185
          | |||  |  | |||||| ||  ||| ||| |||| |||||   |  ||||||  ||||
Db    906 CTAAAAAAAGTGTTCAGTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 965

Qy   1186 AAAAAAAAAAAAA 1199
          |||||||||||||
Db    966 AAAAAAAAAAAAA 979


RESULT  14
US-09-164-193-1
; Sequence 1, Application US/09164193C
; Patent No. 6258582
; GENERAL INFORMATION:
;  APPLICANT: Acton, Susan L.
;  TITLE OF INVENTION: NOVEL CSAPTP NUCLEIC ACID MOLECULES AND USES THEREFOR (as amended
;  FILE REFERENCE: MNI-051
;  CURRENT APPLICATION NUMBER: US/09/164,193C
;  CURRENT FILING DATE: 1998-09-30
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 1315
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (249)..(767)
;   FEATURE:
;   OTHER INFORMATION: 'n' at position 1315 may be any nucleotide
US-09-164-193-1


  Query Match             3.5%;  Score 42;  DB 4;  Length 1315;
  Best Local Similarity   62.3%;  Pred. No. 0.019;
  Matches   66;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy   1094 AGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATA 1153
          ||   ||| |   ||  ||  | ||  | | || || | | |    ||| |  ||  | |
Db   1190 agaccccggggcagtcaggtgctccggacacccgaaggcaataaaacaggagccgtgaaa 1249

Qy   1154 AACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAAAAAAAAA 1199
          || |||| |||||   |  ||||||  ||||||||||||||||||
Db   1250 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1295


RESULT  15
US-08-955-918C-1
; Sequence 1, Application US/08955918C
; Patent No. 6268130
;  GENERAL INFORMATION:
;    APPLICANT:  Kleyn, Patrick, and Moore, Karen
;    TITLE OF INVENTION:  RP Compositions and Therapeutic and
;    TITLE OF INVENTION:  Diagnostic Uses Therefor
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/955,918C
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/697,766
;      FILING DATE:  29-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  MNI-007CPDV2CPA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2184 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  569..1616
US-08-955-918C-1


  Query Match             3.4%;  Score 41.2;  DB 4;  Length 2184;
  Best Local Similarity   54.9%;  Pred. No. 0.044;
  Matches   79;  Conservative    1;  Mismatches   64;  Indels    0;  Gaps    0;

Qy   1056 GATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAAC 1115
          ||   |||| ||   |   ||  | | | || || ||  || | || | ||  ||    |
Db   1981 GAGGAGCTGGGGCGGGCTAGGAGGAGGGGGTAGGTGATGGGGGACGAGGGCCAGGCACCC 2040

Qy   1116 CCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAA 1175
           | |       ||    | |  |  | :    ||| ||| |||| |||||   |  |||
Db   2041 ACATCCCCAATAAAGCCGCGTCCTTGGCMAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2100

Qy   1176 AAACCCAAAAAAAAAAAAAAAAAA 1199
          |||   ||||||||||||||||||
Db   2101 AAAAAAAAAAAAAAAAAAAAAAAA 2124


RESULT  16
US-08-888-077A-41
; Sequence 41, Application US/08888077A
; Patent No. 6020143
;  GENERAL INFORMATION:
;    APPLICANT:  ST. GEORGE-HYSLOP, PETER H
;    APPLICANT:  ROMMENS, JOHANNA M
;    APPLICANT:  FRASER, PAUL E
;    TITLE OF INVENTION:  GENETIC SEQUENCES AND PROTEINS RELATED
;    TITLE OF INVENTION:  TO ALZHEIMER'S DISEASE AND USES THEREFOR.
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LERNER, DAVID, LITTENBERG, KRUMHOLZ & MENTLIK
;      STREET:  600 SOUTH AVENUE WEST
;      CITY:  WESTFIELD
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07090-1497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/888,077A
;      FILING DATE:  03-JUL-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/592,541
;      FILING DATE:  26-JAN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PALISI, THOMAS M
;      REGISTRATION NUMBER:  36,629
;      REFERENCE/DOCKET NUMBER:  SCHERING 3.0-017 CIP CIP CIP IV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908) 654-5000
;      TELEFAX:  (908) 654-7866
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2205 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2205
;      OTHER INFORMATION:  /note= "mutTM1-TM2"
US-08-888-077A-41


  Query Match             3.4%;  Score 41.2;  DB 3;  Length 2205;
  Best Local Similarity   72.2%;  Pred. No. 0.044;
  Matches   52;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy   1128 AAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAA 1187
          |||| | || ||    |||  ||| ||| |||| |||||   |  ||||||   ||||||
Db   2096 AAAGTANTTNTTACACATTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2155

Qy   1188 AAAAAAAAAAAA 1199
          ||||||||||||
Db   2156 AAAAAAAAAAAA 2167


RESULT  17
US-08-361-467B-4
; Sequence 4, Application US/08361467B
; Patent No. 5633441
;  GENERAL INFORMATION:
;    APPLICANT:  De Greef, Willy
;    APPLICANT:  Van Emmelo, John
;    APPLICANT:  De Oliveria, Dulce E.
;    APPLICANT:  De Souza, Maria-Helena
;    APPLICANT:  Van Montagu, Marc
;    TITLE OF INVENTION:  PLANTS WITH MODIFIED FLOWERS, SEEDS OR
;    TITLE OF INVENTION:  EMBRYOS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/361,467B
;      FILING DATE:  22-DEC-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/681,492
;      FILING DATE:  04-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/EP90/01275
;      FILING DATE:  01-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 89 402 224.3
;      FILING DATE:  04-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Schulman, Robert M.
;      REGISTRATION NUMBER:  31,196
;      REFERENCE/DOCKET NUMBER:  010830-027
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1046 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    IMMEDIATE SOURCE:
;      CLONE:  3C9
US-08-361-467B-4


  Query Match             3.4%;  Score 40.6;  DB 1;  Length 1046;
  Best Local Similarity   66.7%;  Pred. No. 0.043;
  Matches   58;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy   1113 AACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATG 1172
          || ||||   ||   || ||| || |    | |||   | ||| |||| |||||   |
Db    928 AAACCCAACAAGTAAAATGGATTTATAAGTTTTTCTTCAAAAAAAAAAAAAAAAAAAAAA 987

Qy   1173 AAAAAACCCAAAAAAAAAAAAAAAAAA 1199
          ||||||   ||||||||||||||||||
Db    988 AAAAAAAAAAAAAAAAAAAAAAAAAA 1014


RESULT  18
US-08-484-332C-4
; Sequence 4, Application US/08484332C
; Patent No. 5767374
;  GENERAL INFORMATION:
;    APPLICANT:  De Greef, Willy
;    APPLICANT:  Van Emmelo, John
;    APPLICANT:  De Oliveria, Dulce E.
;    APPLICANT:  De Souza, Maria-Helena
;    APPLICANT:  Van Montagu, Marc
;    TITLE OF INVENTION:  PLANTS WITH MODIFIED FLOWERS, SEEDS OR
;    TITLE OF INVENTION:  EMBRYOS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,332C
;      FILING DATE:  7-JUNE-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/361,467
;      FILING DATE:  22-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/681,492
;      FILING DATE:  04-APR-1991
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  WO PCT/EP90/01275
;      FILING DATE:  01-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 89 402 224.3
;      FILING DATE:  04-AUG-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Schulman, Robert M.
;      REGISTRATION NUMBER:  31,196
;      REFERENCE/DOCKET NUMBER:  010830-093
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1046 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    IMMEDIATE SOURCE:
;      CLONE:  3C9
US-08-484-332C-4


  Query Match             3.4%;  Score 40.6;  DB 1;  Length 1046;
  Best Local Similarity   66.7%;  Pred. No. 0.043;
  Matches   58;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy   1113 AACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATG 1172
          || ||||   ||   || ||| || |    | |||   | ||| |||| |||||   |
Db    928 AAACCCAACAAGTAAAATGGATTTATAAGTTTTTCTTCAAAAAAAAAAAAAAAAAAAAAA 987

Qy   1173 AAAAAACCCAAAAAAAAAAAAAAAAA 1199
          ||||||   |||||||||||||||||
Db    988 AAAAAAAAAAAAAAAAAAAAAAAAA 1014


RESULT  19
US-08-686-878A-34
; Sequence 34, Application US/08686878A
; Patent No. 5708157
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  LaVallie, Edward
;    APPLICANT:  Racie, Lisa
;    APPLICANT:  Merberg, David
;    APPLICANT:  Treacy, Maurice
;    APPLICANT:  Evans, Cheryl
;    APPLICANT:  Spaulding, Vikki
;    TITLE OF INVENTION:  SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION:  ENCODING THEM
;    NUMBER OF SEQUENCES:  71
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/686,878A
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  216 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-686-878A-34


  Query Match             3.4%;  Score 40.4;  DB 1;  Length 216;
  Best Local Similarity   71.6%;  Pred. No. 0.02;
  Matches   53;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy   1123 AGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCA 1182
          | ||| ||  | |||||  || | | ||| ||| |||| |||||   |  ||||||  |
Db    143 AACCCCAATAAATTTTTGGTTTACTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 202

Qy   1183 AAAAAAAAAAAAAA 1196
          ||||||||||||||
Db    203 AAAAAAAAAAAAAA 216


RESULT  20
US-09-175-928-34
; Sequence 34, Application US/09175928A
; Patent No. 6312921
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT: McCoy, John M.
;  APPLICANT: LaVallie, Edward R.
;  APPLICANT: Collins-Racie, Lisa A.
;  APPLICANT: Evans, Cheryl
;  APPLICANT: Merberg, David
;  APPLICANT: Treacy, Maurice
;  APPLICANT: Mi, Sha
;  APPLICANT: Genetics Institute, Inc.
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: 6006B.AJ172A
;  CURRENT APPLICATION NUMBER: US/09/175,928A
;  CURRENT FILING DATE: 1998-10-20
;  NUMBER OF SEQ ID NOS: 62
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34
;   LENGTH: 216
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (69)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (86)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (114)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (116)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (119)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (142)
US-09-175-928-34


  Query Match             3.4%;  Score 40.4;  DB 4;  Length 216;
  Best Local Similarity   71.6%;  Pred. No. 0.02;
  Matches   53;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;
```

```
Qy   1123 AGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCA 1182
          | ||| ||  | |||||  || | | ||| ||| |||| |||||   |  ||||||   |
Db    143 aaccccaataaatttttggtttactaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 202

Qy   1183 AAAAAAAAAAAAA 1196
          |||||||||||||
Db    203 aaaaaaaaaaaaa 216


RESULT  21
US-09-385-982-280/c
; Sequence 280, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
;  APPLICANT: ENDEGE, WILSON O., ET AL.
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS:  II
;  FILE REFERENCE: CCDNA-260XX
;  CURRENT APPLICATION NUMBER: US/09/385,982
;  CURRENT FILING DATE: 1999-08-30
;  EARLIER APPLICATION NUMBER: 09/328,111
;  EARLIER FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: 60/117,393
;  EARLIER FILING DATE: 1999-01-27
;  EARLIER APPLICATION NUMBER: 60/098,639
;  EARLIER FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 544
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 280
;   LENGTH: 612
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(612)
;   OTHER INFORMATION: n = A,T,C or G
US-09-385-982-280


  Query Match             3.4%;  Score 40.2;  DB 4;  Length 612;
  Best Local Similarity   70.1%;  Pred. No. 0.041;
  Matches   54;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy   1123 AGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCA 1182
          | ||| | || |||  | | | ||  ||| ||| |||| |||||   |  ||| ||   |
Db     80 AACCCCATGGCCTTCATGACTTTTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAA 21

Qy   1183 AAAAAAAAAAAAAAAA 1199
          ||||||||||||||||
Db     20 AAAAAAAAAAAAAAAA 4


RESULT  22
PCT-US96-00994-3
; Sequence 4, Application US/
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  SELF-RENEWING PLURIPOTENT HEMATOPOIETIC
;    TITLE OF INVENTION:  STEM CELL COMPOSITIONS, METHODS OF USE,
;    TITLE OF INVENTION:  AND CULTURE SYSTEMS THEREFOR
;    NUMBER OF SEQUENCES:  4
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: PCT/US96/00994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US SN 08/462,108
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US SN 08/378,144
;      FILING DATE:  24-JAN-1995
; Sequence 3, Application US/
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  SELF-RENEWING PLURIPOTENT HEMATOPOIETIC
;    TITLE OF INVENTION:  STEM CELL COMPOSITIONS, METHODS OF USE,
;    TITLE OF INVENTION:  AND CULTURE SYSTEMS THEREFOR
;    NUMBER OF SEQUENCES:  4
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;    CURRENT APPLICATION DATA:  APPLICATION NUMBER: US UNKNOWN
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US SN 08/462,108
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US SN 08/378,144
;      FILING DATE:  24-JAN-1995
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1721 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  63..1436
PCT-US96-00994-3


  Query Match             3.4%;  Score 40.2;  DB 5;  Length 1721;
  Best Local Similarity   67.1%;  Pred. No. 0.074;
  Matches   57;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy   1115 CCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAA 1174
          | || ||      ||||  |||||   | |||  ||| ||| |||  |||||   |  ||
Db   1627 CACCCCAATCATTAAAGATCTTTTGAGGAATTAAAAAAAAAGAAAGAAAAAAAAAAAAA 1686

Qy   1175 AAAACCCAAAAAAAAAAAAAAAAAA 1199
          ||||   |||||||||||||||||||
Db   1687 AAAAAAAAAAAAAAAAAAAAAAAAA 1711


RESULT  23
US-09-232-191-8
; Sequence 8, Application US/09232191
; Patent No. 6284487
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  TITLE OF INVENTION: Fatty Acid Transport Proteins
;  FILE REFERENCE: WHI97-21p3ME
;  CURRENT APPLICATION NUMBER: US/09/232,191
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8
;   LENGTH: 2301
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-191-8
```

```
  Query Match             3.3%;  Score 40;  DB 4;  Length 2301;
  Best Local Similarity   58.3%;  Pred. No. 0.1;
  Matches   70;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    1080 TGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTT 1139
           ||||||||||| | ||   ||     | | | |   | |    |     |||       ||
Db    2175 tgagtgtgggtggaacaggcatgaggagagtgtggctcaggggccaataaactctgcctt 2234

Qy    1140 CAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAAAAAAAAA 1199
           | |  ||  || ||| |||| |||||   |  ||||||  ||||||||||||||||||
Db    2235 gactcctcttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2294


RESULT  24
US-09-232-200-8
; Sequence 8, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8
;   LENGTH: 2301
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-200-8


  Query Match             3.3%;  Score 40;  DB 4;  Length 2301;
  Best Local Similarity   58.3%;  Pred. No. 0.1;
  Matches   70;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    1080 TGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTT 1139
           ||||||||||| | ||   ||     | | | |   | |    |     |||       ||
Db    2175 tgagtgtgggtggaacaggcatgaggagagtgtggctcaggggccaataaactctgcctt 2234

Qy    1140 CAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAAAAAAAAA 1199
           | |  ||  || ||| |||| |||||   |  ||||||  ||||||||||||||||||
Db    2235 gactcctcttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2294


RESULT  25
US-09-232-197-8
; Sequence 8, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8
;   LENGTH: 2301
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-197-8


  Query Match             3.3%;  Score 40;  DB 4;  Length 2301;
  Best Local Similarity   58.3%;  Pred. No. 0.1;
  Matches   70;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    1080 TGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTT 1139
           ||||||||||| | ||   ||     | | | |   | |    |     |||       ||
Db    2175 tgagtgtgggtggaacaggcatgaggagagtgtggctcaggggccaataaactctgcctt 2234

Qy    1140 CAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAAAAAAAAA 1199
           | |  ||  || ||| |||| |||||   |  ||||||  ||||||||||||||||||
Db    2235 gactcctcttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2294


RESULT  26
US-09-232-201-8
; Sequence 8, Application US/09232201A
; Patent No. 6348321
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MC
;  CURRENT APPLICATION NUMBER: US/09/232,201A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8
;   LENGTH: 2301
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-201-8


  Query Match             3.3%;  Score 40;  DB 4;  Length 2301;
  Best Local Similarity   58.3%;  Pred. No. 0.1;
  Matches   70;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    1080 TGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTT 1139
           ||||||||||| | ||   ||     | | | |   | |    |     |||       ||
Db    2175 tgagtgtgggtggaacaggcatgaggagagtgtggctcaggggccaataaactctgcctt 2234

Qy    1140 CAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAAAAAAAAA 1199
           | |  ||  || ||| |||| |||||   |  ||||||  ||||||||||||||||||
Db    2235 gactcctcttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2294


RESULT  27
5258287-23
;Patent No. 5258287
;    APPLICANT: BAXTER, ROBERT C.;WOOD, WILLIAM I.
;    TITLE OF INVENTION: DNA ENCODING AND METHODS OF PRODUCTION
;OF INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN BP53
```

```
;    NUMBER OF SEQUENCES: 58
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/171,623
;      FILING DATE: 22-MAR-1988
;SEQ ID NO:23:
;      LENGTH: 2550
5258287-23


  Query Match             3.3%;  Score 40;  DB 6;  Length 2550;
  Best Local Similarity   72.2%;  Pred. No. 0.11;
  Matches   52;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy   1128 AAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAA 1187
           |||  |  |||| |  |||  ||| ||| |||| |||||   |  ||||||   ||||||
Db   2457 aaataaagttttaccattaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2516

Qy   1188 AAAAAAAAAAAA 1199
           ||||||||||||
Db   2517 aaaaaaaaaaaa 2528


RESULT  28
5168051-9
;Patent No. 5168051
;    APPLICANT: DERYNCK, RIK M.A.;GOEDDEL, DAVID V.
;    TITLE OF INVENTION: NUCLEIC ACID ENCODING TGF-B ITS USES
;    NUMBER OF SEQUENCES: 21
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/389,929
;      FILING DATE: 04-AUG-1989
;SEQ ID NO:9:
;      LENGTH: 2671
5168051-9


  Query Match             3.3%;  Score 40;  DB 6;  Length 2671;
  Best Local Similarity   72.2%;  Pred. No. 0.11;
  Matches   52;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy   1128 AAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAA 1187
           |||||  |||  |||      |  |||  ||| |||| |||||   |  ||||||   ||||||
Db   2560 aaaggccttattctaaggtgtaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2619

Qy   1188 AAAAAAAAAAAA 1199
           ||||||||||||
Db   2620 aaaaaaaaaaaa 2631


RESULT  29
US-09-232-200-44
; Sequence 44, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 44
;   LENGTH: 2710
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-200-44


  Query Match             3.3%;  Score 40;  DB 4;  Length 2710;
  Best Local Similarity   58.3%;  Pred. No. 0.11;
  Matches   70;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy   1080 TGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTT 1139
           ||||||||||  |  ||   ||     | | | |   | |    |     |||       ||
Db   2584 tgagtgtgggtggaacaggcatgaggagagtgtggctcaggggccaataaactctgcctt 2643

Qy   1140 CAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAAAAAAAAA 1199
           | |  ||  || ||| |||| |||||   |  ||||||   ||||||||||||||||||
Db   2644 gactcctcttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2703


RESULT  30
US-09-232-200-70
; Sequence 70, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 70
;   LENGTH: 2710
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-200-70


  Query Match             3.3%;  Score 40;  DB 4;  Length 2710;
  Best Local Similarity   58.3%;  Pred. No. 0.11;
  Matches   70;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy   1080 TGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTT 1139
           ||||||||||  |  ||   ||     | | | |   | |    |     |||       ||
Db   2584 tgagtgtgggtggaacaggcatgaggagagtgtggctcaggggccaataaactctgcctt 2643

Qy   1140 CAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAAAAAAAAA 1199
           | |  ||  || ||| |||| |||||   |  ||||||   ||||||||||||||||||
Db   2644 gactcctcttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2703


RESULT  31
US-09-232-197-44
; Sequence 44, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
```

```
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 44
;   LENGTH: 2710
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-197-44


  Query Match             3.3%;  Score 40;  DB 4;  Length 2710;
  Best Local Similarity   58.3%;  Pred. No. 0.11;
  Matches   70;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy   1080 TGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTT 1139
           |||||||||| | ||   ||     | | | |   | |    |     |||       ||
Db   2584 tgagtgtgggtggaacaggcatgaggagagtgtggctcaggggccaataaactctgcctt 2643

Qy   1140 CAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAAAAAAAAA 1199
             | |  ||  || ||| |||| |||||   |  ||||||   ||||||||||||||||||
Db   2644 gactcctcttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2703


RESULT  32
US-09-232-197-70
; Sequence 70, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 70
;   LENGTH: 2710
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-197-70


  Query Match             3.3%;  Score 40;  DB 4;  Length 2710;
  Best Local Similarity   58.3%;  Pred. No. 0.11;
  Matches   70;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy   1080 TGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTT 1139
           |||||||||| | ||   ||     | | | |   | |    |     |||       ||
Db   2584 tgagtgtgggtggaacaggcatgaggagagtgtggctcaggggccaataaactctgcctt 2643

Qy   1140 CAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAAAAAAAAA 1199
             | |  ||  || ||| |||| |||||   |  ||||||   ||||||||||||||||||
Db   2644 gactcctcttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2703


RESULT  33
US-09-232-201-44
; Sequence 44, Application US/09232201A
; Patent No. 6348321
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MC
;  CURRENT APPLICATION NUMBER: US/09/232,201A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 44
;   LENGTH: 2710
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-201-44


  Query Match             3.3%;  Score 40;  DB 4;  Length 2710;
  Best Local Similarity   58.3%;  Pred. No. 0.11;
  Matches   70;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy   1080 TGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTT 1139
           |||||||||| | ||   ||     | | | |   | |    |     |||       ||
Db   2584 tgagtgtgggtggaacaggcatgaggagagtgtggctcaggggccaataaactctgcctt 2643

Qy   1140 CAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAAAAAAAAA 1199
             | |  ||  || ||| |||| |||||   |  ||||||   ||||||||||||||||||
Db   2644 gactcctcttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2703


RESULT  34
US-09-232-201-70
; Sequence 70, Application US/09232201A
; Patent No. 6348321
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MC
;  CURRENT APPLICATION NUMBER: US/09/232,201A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 70
;   LENGTH: 2710
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-201-70


  Query Match             3.3%;  Score 40;  DB 4;  Length 2710;
  Best Local Similarity   58.3%;  Pred. No. 0.11;
  Matches   70;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;
```

```
Qy   1080 TGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTT 1139
          |||||||||| | ||   ||     | | | |   | |     |     |||       ||
Db   2584 tgagtgtgggtggaacaggcatgaggagagtgtggctcaggggccaataaactctgcctt 2643

Qy   1140 CAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAAAAAAAAAAAAAA 1199
          | |  ||  || ||| |||| |||||   |  ||||||   ||||||||||||||||||
Db   2644 gactcctcttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2703


RESULT 35
US-08-899-437-1
; Sequence 1, Application US/08899437
; Patent No. 6121415
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
;    TITLE OF INVENTION:  ErbB Receptor-Specific Neuregulin Related
;    TITLE OF INVENTION:  Ligands and Uses Therefor
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/899,437
;      FILING DATE:  24-Jul-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conley, Deirdre L.
;      REGISTRATION NUMBER:  36,487
;      REFERENCE/DOCKET NUMBER:  P1084R1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-2066
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2538 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    FEATURE:
;      NAME/KEY:  mouse NRG3 nucleic acid
;      LOCATION:  1-2538
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
US-08-899-437-1


  Query Match             3.3%;  Score 39.8;  DB 3;  Length 2538;
  Best Local Similarity   49.4%;  Pred. No. 0.12;
  Matches  130;  Conservative    0;  Mismatches  132;  Indels    1;  Gaps    1;

Qy    391 TCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCT 450
          || | |||||   || |     |   |||   |   | |       |  |   ||  |
Db     75 TCCCCTCCCCCGTCCTCTCCTCCCCGCCCGCCGCCCGCCTCTCGGGGGGAGGGGCGTGGG 134

Qy    451 TGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTT 510
          || ||  ||| | ||| |||  || |||  || | ||||| || || ||| | ||
Db    135 GGCAGGGAGCCGATTTGCATGCGGCCGCCGCGGCCGCTGCCTGAGCCGGAGCCCGCCGCC 194

Qy    511 GCCGCTTCCC-CTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTCAGCCAGG 569
          ||||   ||| | |||   ||| | |  | || | || | | | |||| ||   ||
Db    195 GCCGGAGCCCGCGCCCGCGCCCGCGCCCGGCCCGCGCGGCCCCATGCCTCTGGCGCGGCC 254

Qy    570 GGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCC 629
              || || |     |  |   |      ||| ||        | |  |||| | |

Db    255 CTCGGGGGGGCGAAGGTGAAGATCGGCTCCTAGGATGAGTGAAGGGGCGGCCGGTGCCTC 314

Qy    630 AGCACATGGAGCCCCTTGGGCTG 652
              ||| ||| ||| ||| ||| |
Db    315 GCCACCTGGTGCCGCTTCGGCAG 337


RESULT 36
US-09-126-121-1
; Sequence 1, Application US/09126121
; Patent No. 6252051
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
;    TITLE OF INVENTION:  ErbB Receptor-Specific Neuregulin Related
;    TITLE OF INVENTION:  Ligands and Uses Therefor
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/126,121
;      FILING DATE:  30-Jul-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conley, Deirdre L.
;      REGISTRATION NUMBER:  36,487
;      REFERENCE/DOCKET NUMBER:  P1084R1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-2066
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2538 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    FEATURE:
;      NAME/KEY:  mouse NRG3 nucleic acid
;      LOCATION:  1-2538
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
US-09-126-121-1


  Query Match             3.3%;  Score 39.8;  DB 4;  Length 2538;
  Best Local Similarity   49.4%;  Pred. No. 0.12;
  Matches  130;  Conservative    0;  Mismatches  132;  Indels    1;  Gaps    1;

Qy    391 TCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCT 450
          || | |||||   || |     |   |||   |   | |       |  |   ||  |
Db     75 TCCCCTCCCCCGTCCTCTCCTCCCCGCCCGCCGCCCGCCTCTCGGGGGGAGGGGCGTGGG 134

Qy    451 TGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTT 510
          || ||  ||| | ||| |||  || |||  || | ||||| || || ||| | ||
Db    135 GGCAGGGAGCCGATTTGCATGCGGCCGCCGCGGCCGCTGCCTGAGCCGGAGCCCGCCGCC 194

Qy    511 GCCGCTTCCC-CTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTCAGCCAGG 569
          ||||   ||| | |||   ||| | |  | || | || | | | |||| ||   ||
Db    195 GCCGGAGCCCGCGCCCGCGCCCGCGCCCGGCCCGCGCGGCCCCATGCCTCTGGCGCGGCC 254

Qy    570 GGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCC 629
              || || |     |  |   |      ||| ||        | |  |||| | |

Db    255 CTCGGGGGGGCGAAGGTGAAGATCGGCTCCTAGGATGAGTGAAGGGGCGGCCGGTGCCTC 314
```

```
Qy     630 AGCACATGGAGCCCCTTGGGCTG 652
           ||| ||| ||| ||| ||| |
Db     315 GCCACCTGGTGCCGCTTCGGCAG 337


RESULT  37
US-09-385-982-389/c
; Sequence 389, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
;  APPLICANT: ENDEGE, WILSON O., ET AL.
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS:  II
;  FILE REFERENCE: CCDNA-260XX
;  CURRENT APPLICATION NUMBER: US/09/385,982
;  CURRENT FILING DATE: 1999-08-30
;  EARLIER APPLICATION NUMBER: 09/328,111
;  EARLIER FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: 60/117,393
;  EARLIER FILING DATE: 1999-01-27
;  EARLIER APPLICATION NUMBER: 60/098,639
;  EARLIER FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 544
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 389
;   LENGTH: 629
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(629)
;   OTHER INFORMATION: n = A,T,C or G
US-09-385-982-389


  Query Match             3.3%;  Score 39.6;  DB 4;  Length 629;
  Best Local Similarity   65.5%;  Pred. No. 0.062;
  Matches   57;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy    1113 AACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATG 1172
           || |   || |    |||| | |   | || |   | || ||| |||| |||||    |
Db      95 AATCTTTCAAAATNAAAAGTAATGACTTAGAAACTGCAAAAAAAAAAAAAAAAAAAAAAA 36

Qy    1173 AAAAAACCCAAAAAAAAAAAAAAAAA 1199
           ||||||  ||||||||||||||||||||
Db      35 AAAAAAAACAAAAAAAAAAAAAAAAA 9


RESULT  38
US-09-020-956-110
; Sequence 110, Application US/09020956
; Patent No. 6261562
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillin, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
;    NUMBER OF SEQUENCES:  178
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/020,956
;      FILING DATE:  09-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  110:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3410 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-020-956-110


  Query Match             3.3%;  Score 39.6;  DB 4;  Length 3410;
  Best Local Similarity   70.8%;  Pred. No. 0.16;
  Matches   51;  Conservative    1;  Mismatches   20;  Indels    0;  Gaps    0;

Qy    1128 AAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAA 1187
           |||||  || ||   | ||  ||| ||| |||| |||||   |  ||||||  ||||||
Db    3306 AAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3365

Qy    1188 AAAAAAAAAAAA 1199
           |||:||||||||
Db    3366 AAARAAAAAAAA 3377


RESULT  39
US-09-030-607-110
; Sequence 110, Application US/09030607
; Patent No. 6262245
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
;    NUMBER OF SEQUENCES:  224
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/030,607
;      FILING DATE:  25-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  110:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3410 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
```

US-09-030-607-110

Query Match 3.3%; Score 39.6; DB 4; Length 3410;
Best Local Similarity 70.8%; Pred. No. 0.16;
Matches 51; Conservative 1; Mismatches 20; Indels 0; Gaps 0;

```
Qy   1128 AAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAA 1187
          |||||  || ||   | ||  ||| ||| |||| |||||   |  ||||||   ||||||
Db   3306 AAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3365

Qy   1188 AAAAAAAAAAAA 1199
          |||:||||||||
Db   3366 AAARAAAAAAAA 3377
```

RESULT 40
US-09-439-313-110
; Sequence 110, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqui
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 110
; LENGTH: 3410
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-439-313-110

Query Match 3.3%; Score 39.6; DB 4; Length 3410;
Best Local Similarity 70.8%; Pred. No. 0.16;
Matches 51; Conservative 1; Mismatches 20; Indels 0; Gaps 0;

```
Qy   1128 AAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAAAAACCCAAAAAA 1187
          |||||  || ||   | ||  ||| ||| |||| |||||   |  ||||||   ||||||
Db   3306 aaaggctttcttatatgtttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3365

Qy   1188 AAAAAAAAAAAA 1199
          |||:||||||||
Db   3366 aaaraaaaaaaa 3377
```

Search completed: September 12, 2002, 16:08:42
Job time: 19566 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2002, 16:05:23 ; Search time 2588.38 Seconds
(without alignments)
6252.115 Million cell updates/sec

Title: US-09-824-178-2
Perfect score: 1199
Sequence: 1 GTGACAGGAGCCGAAGCAGC.....CCAAAAAAAAAAAAAAAAAA 1199

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	720	60.1	829	10	BI905833	BI905833 603062905
2	709.8	59.2	950	10	BG677538	BG677538 602625181
3	678.4	56.6	1160	10	BM449607	BM449607 AGENCOURT
4	661.2	55.1	968	10	BF348127	BF348127 602022019
5	658	54.9	1014	10	BE791045	BE791045 601586506
6	652.2	54.4	727	10	BI761785	BI761785 603046779
7	641.2	53.5	1092	10	BF312829	BF312829 601896968
8	638.8	53.3	816	10	BG391362	BG391362 602417427
9	631.6	52.7	732	9	AW963045	AW963045 EST375118
10	631	52.6	745	10	BG682778	BG682778 602651269
11	622.6	51.9	729	10	BI906393	BI906393 603063405
12	621.8	51.9	702	10	BI225153	BI225153 602949826
13	620.6	51.8	704	10	BG574918	BG574918 602598004
14	617.6	51.5	901	10	BE799496	BE799496 601589290
15	613.4	51.2	765	10	BI753979	BI753979 603027411
16	608.4	50.7	703	10	BG742185	BG742185 602634070
17	605	50.5	927	10	BI195920	BI195920 602756190
18	591.2	49.3	1140	10	BF981595	BF981595 602309483
19	589.2	49.1	1139	11	AK006149	AK006149 Mus muscu
20	585.6	48.8	824	10	BI222786	BI222786 602941568
21	580.8	48.4	606	10	BE906167	BE906167 601497230
22	577.8	48.2	688	10	BG741487	BG741487 602632285
23	577.6	48.2	972	10	BF795934	BF795934 602259188
24	577.4	48.2	1097	10	BM478401	BM478401 AGENCOURT
25	575.8	48.0	731	10	BG716405	BG716405 602676303
26	574.8	47.9	670	10	BI835836	BI835836 603087830
27	574.4	47.9	713	10	BI223805	BI223805 602942709
28	571.6	47.7	975	10	BG180105	BG180105 602329682
29	570.8	47.6	808	9	AU137227	AU137227 AU137227
30	566.4	47.2	647	10	BE882748	BE882748 601505521
31	566	47.2	719	10	BI521459	BI521459 603081263
32	560.6	46.8	729	9	AV727450	AV727450 AV727450
33	558.4	46.6	659	10	BI224994	BI224994 602949185
34	556	46.4	638	10	BG248926	BG248926 602361363
35	550.8	45.9	682	10	BG748595	BG748595 602706289
36	550.6	45.9	614	10	BI255306	BI255306 602977333
37	549.8	45.9	636	9	AW965834	AW965834 EST377907
38	546.4	45.6	983	10	BF982805	BF982805 602304840
39	540.6	45.1	707	10	BG576751	BG576751 602599070
40	535.2	44.6	718	10	BM018285	BM018285 603645940
41	535	44.6	614	10	BG290416	BG290416 602388262
42	533.6	44.5	746	10	BG290761	BG290761 602389081
43	532.4	44.4	689	10	BE387277	BE387277 601275467
44	529.2	44.1	671	10	BI833929	BI833929 603085025
c 45	524.6	43.8	633	10	BF220007	BF220007 601296902

ALIGNMENTS

RESULT 1
BI905833
LOCUS BI905833 829 bp mRNA linear EST 16-OCT-2001
DEFINITION 603062905F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5212129 5', mRNA sequence.
ACCESSION BI905833
VERSION BI905833.1 GI:16168424
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 829)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11532 row: j column: 02
High quality sequence start: 2
High quality sequence stop: 815.
FEATURES Location/Qualifiers
source 1. .829
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5212129"
/clone_lib="NIH_MGC_118"
/tissue_type="leukocyte"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source leukocytes from anonymous pool of non-activated adult donors. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon

cloning). Average insert size 1.7 kb, insert size range 1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH_MGC Library."
BASE COUNT 205 a 223 c 234 g 167 t
ORIGIN

Query Match 60.1%; Score 720; DB 10; Length 829;
Best Local Similarity 95.0%; Pred. No. 1.3e-107;
Matches 796; Conservative 0; Mismatches 15; Indels 27; Gaps 4;

```
Qy      33 CCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCCCTTACACTCCACA 92
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCCCTTACACTCCACA 60

Qy      93 GTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCG 152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCG 120

Qy     153 AGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAA 212
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAA 180

Qy     213 AGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGA 272
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGA 240

Qy     273 AGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCA 332
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCA 300

Qy     333 AGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATC 392
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATC 360

Qy     393 ACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTG 452
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTG 420

Qy     453 CGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTTGC 512
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGA---------- 470

Qy     513 CGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTT----------C 562
           ||||||||||||||||||||||||||||||||||||||||||||||||||          |
Db     471 CGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGC 530

Qy     563 AGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTG 622
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     531 AGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTG 590

Qy     623 GTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTT 682
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     591 GTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTT 650

Qy     683 ACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGG----GGTGGGGG 738
           ||||||||||||||||||||||||||||||||||||||||||||||||    || ||| |
Db     651 ACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGAGTGGAGGGTG 710

Qy     739 TGGGGCTGAAAGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAGGGCTTGGAGTTGG- 797
           |||     |||||||| ||||||||||||||||||||||||||||||||||  | |
Db     711 GGGGTTGACAAGGGTGGAGTGGGAGATATGGAGGAGTATGAATTAGGGCTTGGACGTTGG 770

Qy     798 --TAAAAACATTCCTGACTATCCTTCTTAACCACGTGGCTGATGTGGGGTAGGTATGA 853
             ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     771 TAAAAAACAATTCCTGACTATCCTTCTTAACCACGTGGCTGATGTGGGGTAGGTATGA 828
```

RESULT 2
BG677538
LOCUS BG677538 950 bp mRNA linear EST 01-MAY-2001

DEFINITION 602625181F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4749981 5', mRNA sequence.
ACCESSION BG677538
VERSION BG677538.1 GI:13908935
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 950)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10604 row: a column: 22
High quality sequence stop: 895.
FEATURES Location/Qualifiers
source 1. .950
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4749981"
/clone_lib="NCI_CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 228 a 248 c 277 g 197 t
ORIGIN

Query Match 59.2%; Score 709.8; DB 10; Length 950;
Best Local Similarity 94.3%; Pred. No. 5.7e-106;
Matches 874; Conservative 0; Mismatches 22; Indels 31; Gaps 12;

```
Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAA-G 179
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAACG 180

Qy     180 AAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAG 239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAG 240

Qy     240 GACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACT 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACT 300

Qy     300 TTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTG 359
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTG 360

Qy     360 CAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCC 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCC 420

Qy     420 AGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCC 479
```

```
Db    421 AGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCC 480

Qy    480 CGGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGT 539
          ||||||||||||||||||||||          ||||||||||||||||||||||||||||
Db    481 CGGGGCTCTGCCAGATCCTGAGA----------CGCTTCCCCTCCCTGCCCCACCCGGGT 530

Qy    540 CCTGTGCTGGCTCCTGCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCT-CGGG 588
          ||||||||||||||||||||||          |||||||||||||||||||||| ||||
Db    531 CCTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCCGGG 590

Qy    589 TGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGG 648
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    591 TGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGT-CCAGCACATGGAGCCCCTTGG 649

Qy    649 GCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAA-TAACAGTTGGG 707
          |||||||||| ||||||||||||||||||||||||||||||||||||| ||||||||||
Db    650 GCTGAGCACC-AGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATTAACAGTTGGG 708

Qy    708 AGAAATATCAATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATAT 767
          ||| |||||| |||||||||||||||| |||||||| |||||||||||   |      |
Db    709 AGACATATCACTGAAATTCTGGGGGTGAGGGTGGGGTTGAAAGGGTGGGTGCGAGATACT 768

Qy    768 GGAGGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACA-TTCCTGACTATCCTTCTTAAC 826
          |||||||||||| |||||||||||||||||||||||| ||||||| ||||||||||||
Db    769 GGAGGAGTATGACTTAGGGCTTGGAGTTGGTAAAAACATTTCCTGATTATCCTTCTTAAC 828

Qy    827 CAC-GTGGCTGATGTGGGGT-AGGTATGAGGGGAA-GGAAGTGGAGTAGCCTAATGAAAA 883
          ||| ||||||||||||||| |||||||| ||||| |||||||||||||||||||||||
Db    829 CACGGTGGCTGATGTGGGGTAAGGTATGACGGGAACGGAAGTGGAGTAGCCTAATGAAA- 887

Qy    884 GGGGTTCTAGTTGAGCTCTGTAGATAA 910
          ||| |   ||||||||||||||||
Db    888 -GGGGTTCTAGTGAGCTCTGTAGATAA 913
```

RESULT 3
BM449607
LOCUS BM449607 1160 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6386356 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5526789 5', mRNA sequence.
ACCESSION BM449607
VERSION BM449607.1 GI:18498647
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1160)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12200 row: p column: 22
High quality sequence stop: 705.

FEATURES Location/Qualifiers
source 1. .1160
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5526789"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."
BASE COUNT 366 a 214 c 356 g 144 t. 80 others
ORIGIN

Query Match 56.6%; Score 678.4; DB 10; Length 1160;
Best Local Similarity 94.8%; Pred. No. 6.7e-101;
Matches 728; Conservative 0; Mismatches 20; Indels 20; Gaps 2;

```
Qy      6 AGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTT 65
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTT 60

Qy     66 GCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAG 125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAG 120

Qy    126 AAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTA 185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTA 180

Qy    186 TTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAA 245
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAA 240

Qy    246 AGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACT 305
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACT 300

Qy    306 CAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGAC 365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGAC 360

Qy    366 CAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAA 425
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAA 420

Qy    426 AGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGC 485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGC 480

Qy    486 TCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTG 545
          ||||||||||||||||          ||||||||||||||||||||||||||||||||||
Db    481 TCTGCCAGATCCTGAGA----------CGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTG 530

Qy    546 CTGGCTCCTGCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGC 595
          ||||||||||||||||          ||||||||||||||||||||||||||||||||||
Db    531 CTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGC 590

Qy    596 TGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGC 655
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    591 TGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGC 650

Qy    656 ACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATAT 715
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    651 ACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATAT 710

Qy    716 CAATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAG 763
          |||||||||| ||| |  ||   ||| | |||    ||  ||| |
Db    711 CAATGAAATTTGGGGTGGNGGNTGGGGNTTTAAAAAAGAGGTGGGGNG 758
```

RESULT 4
BF348127
LOCUS BF348127 968 bp mRNA linear EST 22-NOV-2000
DEFINITION 602022019F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4157548 5', mRNA sequence.
ACCESSION BF348127
VERSION BF348127.1 GI:11295722
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 968)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9432  row: e  column: 05
            High quality sequence stop: 725.
FEATURES             Location/Qualifiers
     source          1. .968
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4157548"
                     /clone_lib="NCI_CGAP_Brn67"
                     /tissue_type="anaplastic oligodendroglioma with 1p/19q
                     loss"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.3 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      237 a    245 c    285 g    201 t
ORIGIN
```

Query Match 55.1%; Score 661.2; DB 10; Length 968;
Best Local Similarity 93.2%; Pred. No. 4.4e-98;
Matches 777; Conservative 0; Mismatches 28; Indels 29; Gaps 7;

```
Qy    299 TTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGT 358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGT 60

Qy    359 GCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCC 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCC 120

Qy    419 CAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGC 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGC 180

Qy    479 CCGGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGG 538
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCGGGGCTCTGCCAGATCCTGAGA----------CGCTTCCCCTCCCTGCCCCACCCGGG 230

Qy    539 TCCTGTGCTGGCTCCTGCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGG 588
          |||||||||||||||||||||||          |||||||||||||||||||||||||||||
Db    231 TCCTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGG 290

Qy    589 TGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGG 648
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 TGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGG 350

Qy    649 GCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGA 708
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 GCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGA 410

Qy    709 GAAATATCAATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATG 768
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    411 GAAATATCAATGAAATTCTGGGGGTGGGGGTGGGGTTGAAAGGGTGGGGTGGGAGATATG 470

Qy    769 GAGGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCA 828
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    471 GAGGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCA 530

Qy    829 CGTGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGT 888
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    531 CGTGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGT 590

Qy    889 TCTAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGA 948
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||| ||
Db    591 TCTAGTTGAGCTCTGTAGAT-AATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCCGA 649

Qy    949 TAAAAGAAACTCCATCCGCACAGACAGATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAG 1008
          | ||||| |||||||||||||| ||| ||||| ||||||||||||||||||||||||||||
Db    650 TTAAAGACACTCCATCCGCACCGACCGATGC-AACAGCTCCTCTAGTTCTGCAGAGCTAG 708

Qy   1009 TTGAGAACTCAACA-TTAATCATTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGG 1067
          |||||||||| || |||||||||||||| |||||||| | |  |||||| |||||||||
Db    709 GTGAGAACTCACCATTTAATCATTTTAACAAGTACTG--CCTTGAATAGACTTGCTGTGG 766

Qy   1068 GAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAACCCCACA 1121
          || |||| |       ||| | |||| |  ||  |||||||||| ||||| |
Db    767 GAGAAAGGCC----GCGGTGTGGGAAACGGACCCCGAGCGTGGGGACCCCCAGA 816
```

```
RESULT   5
BE791045
LOCUS       BE791045                1014 bp    mRNA    linear   EST 20-SEP-2000
DEFINITION  601586506F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3940752 5',
            mRNA sequence.
ACCESSION   BE791045
VERSION     BE791045.1  GI:10212243
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1014)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM791  row: l  column: 01
            High quality sequence stop: 615.
FEATURES             Location/Qualifiers
     source          1. .1014
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3940752"
                     /clone_lib="NIH_MGC_7"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      271 a    243 c    290 g    210 t
ORIGIN
```

Query Match 54.9%; Score 658; DB 10; Length 1014;
Best Local Similarity 94.6%; Pred. No. 1.4e-97;
Matches 732; Conservative 0; Mismatches 20; Indels 22; Gaps 4;

```
Qy      26 AGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCCCTTACA 85
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCCCTTACA 60

Qy      86 CTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAG 145
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAG 120

Qy     146 ACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAA 205
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAA 180

Qy     206 GAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTC 265
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTC 240

Qy     266 CTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCC 325
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCC 300

Qy     326 AAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACT 385
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACT 360

Qy     386 GGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGC 445
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGC 420

Qy     446 AAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTG 505
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGA--- 477

Qy     506 CTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTT---- 561
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     478 -------CGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTCCTG 530

Qy     562 ------CAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTT 615
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     531 CTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTT 590

Qy     616 CCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTT 675
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     591 CCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACTTTTG 650

Qy     676 TTGTTTTACCTTTTTTCCAAAT-AACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTG 734
           |  |||  | ||||| | |||| |||||||||||| || ||||||| || |||||||||
Db     651 TGTTTTACCTTTTTTCCAAAATAAACAGTTGGGAGACATTTCAATGACATACTGGGGGTG 710

Qy     735 GGGGTGGGG-CTGAAAGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAGGGC 787
           |||||||||  |||| |||||||||||||||||||| |||||||| |  || |
Db     711 GGGGTGGGGTTTGAACGGGTGGGGTGGGAGATATGGACGAGTATGACTAGGGTC 764
```

```
RESULT   6
BI761785
LOCUS       BI761785                727 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603046779F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5187258 5',
            mRNA sequence.
ACCESSION   BI761785
VERSION     BI761785.1  GI:15753363
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 727)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11467  row: m  column: 19
            High quality sequence stop: 727.
FEATURES             Location/Qualifiers
     source          1. .727
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5187258"
                     /clone_lib="NIH_MGC_116"
                     /lab_host="DH10B"
                     /note="Organ: pooled colon, kidney, stomach; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of 3 colons, age 26 yo male, 49 yo
                     female, 71 yo male colon; 46 yo male kidney, and pool of 2
                     stomachs, 62 yo male and 70 yo female.  Library is
                     oligo-dT primed and  directionally cloned (EcoRV site is
                     destroyed upon cloning).  Average insert size 1.4 kb,
                     insert size range 1-3 kb.  Library is normalized and
                     enriched for full-length clones and was constructed by C.
                     Gruber  (Invitrogen).  Research Genetics tracking code
                     023. Note: this is a NIH_MGC Library."
BASE COUNT      173 a    214 c    199 g    140 t      1 others
ORIGIN
```

```
  Query Match             54.4%;  Score 652.2;  DB 10;  Length 727;
  Best Local Similarity   96.7%;  Pred. No. 1.4e-96;
  Matches  694;  Conservative    0;  Mismatches    4;  Indels   20;  Gaps    2;

Qy      33 CCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCCCTTACACTCCACA 92
           |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20 CCCGNTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCCCTTACACTCCACA 79

Qy      93 GTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCG 152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80 GTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCG 139

Qy     153 AGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAA 212
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     140 AGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAA 199

Qy     213 AGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGA 272
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     200 AGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGA 259

Qy     273 AGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCA 332
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     260 AGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCA 319

Qy     333 AGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATC 392
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     320 AGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATC 379

Qy     393 ACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTG 452
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     380 ACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTG 439

Qy     453 CGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTTGC 512
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     440 CGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGA---------- 489

Qy     513 CGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTT----------C 562
           ||||||||||||||||||||||||||||||||||||||||||||||||||          |
Db     490 CGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGC 549

Qy     563 AGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTG 622
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     550 AGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTG 609
```

```
Qy     623 GTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTT 682
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     610 GTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTT 669

Qy     683 ACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGTG 740
           |||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db     670 ACCTTTTTTCCAAATAACAGTTGGGAGAAATATTCATGAAATTCTGGGGGTGGGGGTG 727


RESULT   7
BF312829
LOCUS       BF312829               1092 bp    mRNA    linear   EST 21-NOV-2000
DEFINITION  601896968F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4126144 5',
            mRNA sequence.
ACCESSION   BF312829
VERSION     BF312829.1  GI:11260709
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1092)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1015  row: h  column: 17
            High quality sequence stop: 661.
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4126144"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      382 a    268 c    293 g    148 t      1 others
ORIGIN


  Query Match           53.5%;  Score 641.2;  DB 10;  Length 1092;
  Best Local Similarity  89.7%;  Pred. No. 7.5e-95;
  Matches  730;  Conservative    0;  Mismatches   63;  Indels   21;  Gaps    3;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 64

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      65 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 124

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     125 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 184

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     185 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 244

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     245 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 304

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     305 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 364

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     365 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 424

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     425 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 484

Qy     481 GGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTC 540
           |||||||||||||||||||||          ||||||||||||||| ||||||||||||
Db     485 GGGGCTCTGCCAGATCCTGAGA----------CGCTTCCCCTCCCTGGCCCACCCGGGTC 534

Qy     541 CTGTGCTGGCTCCTGCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTG 590
           ||||||||||||||||||||          ||| ||||||||||||||||||||||||||
Db     535 CTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGGCAGGGGTCAGGAGGTGGCTCGGGTG 594

Qy     591 T-GGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGG 649
           | ||||| ||||||||||||||||||||||||||||||||| |||||||||| | |||
Db     595 TGGGGCTTGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGGACATGGAGCCCATGGGG 654

Qy     650 CTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAG 709
           |||||||||||||||||||||||||    | || |||||   |||||||  |  |||| ||
Db     655 CTGAGCACCAAGACCTTGAACCTTTGAAGGTTGACCTTGAATCCAAATTAAGGTTGCGAA 714

Qy     710 AAATATCAATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGG 769
           ||| ||    |||||   ||| |||| ||||||    ||  || ||    | | ||| |
Db     715 AAAAATATCAGAAATCTGGGGAGTGGTGGTGGGATAAAAGAGGGGGATGAGAAAATAGGA 774

Qy     770 AGGAGTATGAATTAGGGCTTGGAGTTGGTAAAAA 803
           ||      |   | ||   |||| ||| | |
Db     775 GAGATATAACAAAGGAGCAGAGAGTGAGTACACA 808


RESULT   8
BG391362
LOCUS       BG391362               816 bp    mRNA    linear   EST 12-MAR-2001
DEFINITION  602417427F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4536942 5',
            mRNA sequence.
ACCESSION   BG391362
VERSION     BG391362.1  GI:13284810
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 816)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10461  row: e  column: 07
            High quality sequence stop: 709.
FEATURES             Location/Qualifiers
     source          1. .816
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4536942"
                     /clone_lib="NIH_MGC_92"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      210 a    185 c    232 g    189 t
ORIGIN


  Query Match             53.3%;  Score 638.8;  DB 10;  Length 816;
  Best Local Similarity   91.3%;  Pred. No. 2e-94;
  Matches  741;  Conservative    0;  Mismatches   47;  Indels   24;  Gaps    5;

Qy    390 ATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGC 449
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     17 ATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGC 76

Qy    450 TTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTT 509
          |||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     77 TTGCGGGTGGCCAAGTTGAATGATGCTGCCCAGGGCTCTGCCAGATCCTGAGA------- 129

Qy    510 TGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTT-------- 561
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    130 ---CGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTCCTGCTTT 186

Qy    562 --CAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTG 619
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    187 TGCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTG 246

Qy    620 TTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGT 679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    247 TTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGT 306

Qy    680 TTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGT 739
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    307 TTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGT 366

Qy    740 GGGGCTGAAAGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAGGGCTTGGAGTTGGTA 799
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    367 GGGGTTGAAAGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAGGGCTTGGAGTTGGTA 426

Qy    800 AAAACATTCCTGACTATCCTTCTTAACCACGTGGCTGATGTGGGGTAGGTATGAGGGGAA 859
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    427 AAAACATTCCTGACTATCCTTCTTAACCACGTGGCTGATGTGGGGTAGGTATGAGGGGAA 486

Qy    860 GGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTAGTTGAGCTCTGTAGATAAATGCCTTGT 919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    487 GGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTAGTTGAGCTCTGTAGATAAATGCCTTGT 546

Qy    920 TTCAGTGTGGTTGGAGACCTGGTGTCAGATAAAAGAAACTCCATCCGCACAGACAGATGC 979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    547 TTCAGTGTGGTTGGAGACCTGGTGTCAGATAAAAGAAACTCCATCCGCACAGACAGATGC 606

Qy    980 AAACAGCTCCTCTAGTTCTGCAGAGCTAGTTGAGAACTCAACATTAA-TCATTTTAAAAA 1038
          ||||||||||||||||||||||||||||||||||||||||||||||| |||||| | |||
Db    607 AAACAGCTCCTCTAGTTCTGCAGAGCTAGTTGAGAACTCAACATTAACTCATTTCACAAA 666

Qy   1039 GTACTGTCCTTGA-AATAGATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGA 1097
          |||||||||||| |||||||||||||||||||| |||||||| ||||||| ||  |
Db    667 GTACTGTCCTTGACAATAGATTTGCTGTGGGAAGAAAGGCAGTGAATGTGGGACAACAGG 726

Qy   1098 GCCGTGAGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACA 1157
           |||||||||||||  | |||| |||||||| |   ||    |       ||  | | |
Db    727 ACCGTGAGCGTGGG--AACCCAAAGAGCCCAAGGACTTTCAGAATCGAAATAACAACAAA 784

Qy   1158 AAACAAAAACCCATGAAAAAACCCAAAAAAAA 1189
          |  |  ||| |||   | ||    ||| || |
Db    785 AGCCTGAAAACCAACCAGAATTTTAAACAAGA 816
```

```
RESULT   9
AW963045
LOCUS       AW963045                732 bp    mRNA    linear   EST 01-JUN-2000
DEFINITION  EST375118 MAGE resequences, MAGH Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW963045
VERSION     AW963045.1  GI:8152881
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 732)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 183
            Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source          1. .732
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGH"
                     /note="Vector: pBluescriptSKm"
BASE COUNT      175 a    210 c    201 g    141 t      5 others
ORIGIN


  Query Match             52.7%;  Score 631.6;  DB 9;  Length 732;
  Best Local Similarity   94.3%;  Pred. No. 3.1e-93;
  Matches  694;  Conservative    0;  Mismatches   21;  Indels   21;  Gaps    3;

Qy     18 AGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGC 77
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGC 60

Qy     78 CCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTG 137
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTG 120

Qy    138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 180

Qy    198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 240

Qy    258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 300

Qy    318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 360

Qy    378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 420

Qy    438 TCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCC 497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCC 480
```

```
Qy     498 TGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCC 557
           |||||          ||||||||||||||||||||||||||||||||||||||||||||
Db     481 TGAGA----------CGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCC 530

Qy     558 CCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGA 607
           ||||          ||||||||||||||||||||||||||||||||||||||||||| |
Db     531 CCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCANA 590

Qy     608 AGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTG 667
           | ||||||  |||||||||||||||||||  ||||||||||||||||||||||||||||
Db     591 ANCCCTTTTCTGTTGGTGTCCCAGCACATTAAGCCCCTTGGGCTGAGCACCAAGACCTTG 650

Qy     668 AACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCT 727
            |||||||||| |||||||||||| |||||||||||| |||  || || |||| || |
Db     651 -ACCTTTTTTGGTTTACCTTTTTTTCAAATAACAGTTTGGAAGAAAATTAATGGAAATTC 709

Qy     728 GGGGGTGGGGGTGGGG 743
             ||| || || ||||
Db     710 CTGGGGGGTGGGGGGG 725
```

```
RESULT  10
BG682778
LOCUS       BG682778                745 bp    mRNA    linear   EST 01-MAY-2001
DEFINITION  602651269F1 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:4762097 5',
            mRNA sequence.
ACCESSION   BG682778
VERSION     BG682778.1  GI:13914175
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 745)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1615  row: j  column: 18
            High quality sequence stop: 698.
FEATURES             Location/Qualifiers
     source          1. .745
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4762097"
                     /clone_lib="NIH_MGC_47"
                     /tissue_type="neuroblastoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      191 a    222 c    200 g    132 t
ORIGIN
```

```
  Query Match             52.6%;  Score 631;  DB 10;  Length 745;
  Best Local Similarity   95.0%;  Pred. No. 3.8e-93;
  Matches  691;  Conservative    0;  Mismatches   15;  Indels   21;  Gaps    3;
```

```
Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      29 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 88

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      89 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 148

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     149 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 208

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     209 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 268

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     269 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 328

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     329 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 388

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     389 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 448

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     449 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 508

Qy     481 GGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTC 540
           ||||||||||||||||||||          ||||||||||||||||||||||||||||||
Db     509 GGGGCTCTGCCAGATCCTGAGA----------CGCTTCCCCTCCCTGCCCCACCCGGGTC 558

Qy     541 CTGTGCTGGCTCCTGCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTG 590
           ||||||||||||||||||||          ||||||||||||||||||||||||||||||
Db     559 CTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTG 618

Qy     591 TGGGCTGGAGAGGCAGAAGCCC-TTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGG 649
           |||||||||||||||||||||| |||||||| |    |||| ||||||||||||||||||
Db     619 TGGGCTGGAGAGGCAGAAGCCCATTTCCTGTTTGGTGTCCAGAACATGGAGCCCCTTGGG 678

Qy     650 CTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAG 709
           |||||||||||||||||||||| |    |    ||||| ||||||||||||||||||||||
Db     679 CTGAGCACCAAGACCTTGAACCTATAAAGAAAAACCTTATTTCCAAATAACAGTTGGGAG 738

Qy     710 AAATATC 716
           |||||||
Db     739 AAATATC 745
```

```
RESULT  11
BI906393
LOCUS       BI906393                729 bp    mRNA    linear   EST 16-OCT-2001
DEFINITION  603063405F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5222512 5',
            mRNA sequence.
ACCESSION   BI906393
VERSION     BI906393.1  GI:16169075
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 729)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
```

```
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM11533  row: j  column: 01
           High quality sequence stop: 681.
FEATURES             Location/Qualifiers
     source          1. .729
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5212512"
                     /clone_lib="NIH_MGC_118"
                     /tissue_type="leukocyte"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV
                     (destroyed); RNA source leukocytes from anonymous pool of
                     non-activated  adult donors.  Library is oligo-dT primed
                     and directionally cloned  (EcoRV site is destroyed upon
                     cloning).  Average insert size 1.7 kb, insert size range
                     1.2-3.3 kb.  Library is normalized and enriched for
                     full-length clones and was constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking code 027. Note:
                     this is a NIH_MGC Library."
BASE COUNT      182 a    221 c    199 g    127 t
ORIGIN


  Query Match             51.9%;  Score 622.6;  DB 10;  Length 729;
  Best Local Similarity   94.5%;  Pred. No. 9e-92;
  Matches  685;  Conservative    0;  Mismatches   19;  Indels   21;  Gaps    3;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 70

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71 CGG-TGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 129

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     130 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 189

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     190 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 249

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     250 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 309

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     310 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 369

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     370 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 429

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     430 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 489

Qy     481 GGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTC 540
           ||||||||||||||||||||          ||||||||||||||| ||||||||||||
Db     490 GGGGCTCTGCCAGATCCTGAGA----------CGCTTCCCCTCCCTGGCCCACCCGGGTC 539

Qy     541 CTGTGCTGGCTCCTGCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTG 590
           ||||||||||||||||||||          ||||||||||||||||||||||||||||||
Db     540 CTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTG 599

Qy     591 TGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGC 650
           ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     600 TGGGCTGGAGAGGCAGAAGCCCTTTCCTGTAGGTGTCCCAGCACATGGAGCCCCTTGGGC 659

Qy     651 TGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGA 710
           |||||||||||||||||| ||| |  || |  |  |||| | | ||||||||||||
Db     660 TGAGCACCAAGACCTTGAAGCTTATACGTATACCTGTTTTCCAACATAACAGTTGGGGAG 719

Qy     711 AATAT 715
             | ||
Db     720 GAAAT 724


RESULT  12
BI225153
LOCUS       BI225153                 702 bp    mRNA    linear   EST 11-JUL-2001
DEFINITION  602949826F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:5094079 5',
            mRNA sequence.
ACCESSION   BI225153
VERSION     BI225153.1  GI:14678597
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 702)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLCM1862  row: c  column: 08
           High quality sequence stop: 695.
FEATURES             Location/Qualifiers
     source          1. .702
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5094079"
                     /clone_lib="NIH_MGC_8"
                     /tissue_type="Burkitt lymphoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      168 a    160 c    211 g    163 t
ORIGIN


  Query Match             51.9%;  Score 621.8;  DB 10;  Length 702;
  Best Local Similarity   95.4%;  Pred. No. 1.2e-91;
  Matches  680;  Conservative    0;  Mismatches   12;  Indels   21;  Gaps    3;

Qy     255 GCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACT 314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACT 60

Qy     315 ACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGA 374
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGA 120

Qy     375 ACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGC 434
```

```
Db    121 ACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGC 180

Qy    435 TCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGA 494

Db    181 TCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGA 240

Qy    495 TCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCT 554

Db    241 TCCTGAGA----------CGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCT 290

Qy    555 GCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGC 604

Db    291 GCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGC 350

Qy    605 AGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACC 664

Db    351 AGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACC 410

Qy    665 TTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAAT 724

Db    411 TTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAAT 470

Qy    725 TCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAG 784

Db    471 TCTGGGGGTGGGGGTGGGGTTGAAAGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAG 530

Qy    785 GGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACGTGGCTGATGTGGGG 844

Db    531 GGCTTGGACGTGGTAAAAACATTCCTGACTATCCTTCTTAACCACGTGGCTGATGTGGGG 590

Qy    845 TAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTAGTTGAGCTCTGT 904

Db    591 TAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTAGTTGAGCTCTGT 650

Qy    905 AGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATAAAAGAAA 957

Db    651 AGAT-AATGCCTTGTTTCAGTTGTGGTTGGAGACTGGTGTCAGATAAAAGAAA 702
```

RESULT 13
BG574918
LOCUS BG574918 704 bp mRNA linear EST 10-APR-2001
DEFINITION 602598004F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4706737 5', mRNA sequence.
ACCESSION BG574918
VERSION BG574918.1 GI:13582571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 704)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10575 row: h column: 02
High quality sequence stop: 665.
FEATURES Location/Qualifiers
source 1. .704
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4706737"
/clone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.383 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 174 a 213 c 191 g 126 t
ORIGIN

Query Match 51.8%; Score 620.6; DB 10; Length 704;
Best Local Similarity 95.1%; Pred. No. 1.9e-91;
Matches 680; Conservative 0; Mismatches 14; Indels 21; Gaps 3;

```
Qy      8 GAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGC 67

Db      1 GAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGC 60

Qy     68 CTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAA 127

Db     61 CTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAA 120

Qy    128 GAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATT 187

Db    121 GAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATT 180

Qy    188 CTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAG 247

Db    181 CTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAG 240

Qy    248 CCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCA 307

Db    241 CCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCA 300

Qy    308 GGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCA 367

Db    301 GGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCA 360

Qy    368 GACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAG 427

Db    361 GACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAG 420

Qy .  428 TCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTC 487

Db    421 TCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTC 480

Qy    488 TGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCT 547

Db    481 TGCCAGATCCTGAGA----------CGCTTCCCCTCCCTGCCCCAACCGGGTCCTGTGCT 530

Qy    548 GGCTCCTGCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTG 597

Db    531 GGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTG 590

Qy    598 GAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCAC 657

Db    591 GAGAGGCAGAAGCCCTTTCCTGTAGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCAC 650

Qy    658 CAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAA 712

Db    651 CAAGACCTTGAACC-TTAAGGTTTACCTTTTTTCCAAATAATCAGTTGGGAGACA 704
```

RESULT 14
BE799496
LOCUS BE799496 901 bp mRNA linear EST 20-SEP-2000
DEFINITION 601589290F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3943271 5', mRNA sequence.
ACCESSION BE799496
VERSION BE799496.1 GI:10220694
KEYWORDS EST.
SOURCE human.

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 901)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM798  row: d  column: 24
            High quality sequence start: 36
            High quality sequence stop: 780.
FEATURES             Location/Qualifiers
     source          1. .901
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3943271"
                     /clone_lib="NIH_MGC_7"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      228 a    196 c    276 g    201 t
ORIGIN
```

```
  Query Match           51.5%;  Score 617.6;  DB 10;  Length 901;
  Best Local Similarity  89.9%;  Pred. No. 5.4e-91;
  Matches  719;  Conservative    0;  Mismatches   74;  Indels    7;  Gaps    5;

Qy    162 AGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGG 221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     92 AGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGG 151

Qy    222 CCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCC 281
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    152 CCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCC 211

Qy    282 AGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGA 341
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    212 AGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGA 271

Qy    342 ATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCA 401
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    272 ATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCA 331

Qy    402 CCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCC 461
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    332 CCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCC 391

Qy    462 AAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCC 521
          |||||||||||||||||||||||||||||||||||||||||| ||||   |  | | |||
Db    392 AAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGAC-GCTTCCCCTCCCTGCCC 450

Qy    522 TCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTCAGCCAGGGGTCAGGAGGTG 581
            || |  ||      | |   |||    ||||||   | ||||||||||||||||||||
Db    451 CACCGGGTCCTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTG 510

Qy    582 GCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGC 641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    511 GCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGC 570

Qy    642 CCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACA 701
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    571 CCCTTGGGCTGAGCACCAAGACCTTGAACC-TTTTTGTTTTACCTTTTTTCCAAATAACA 629

Qy    702 GTTGGGAGAAATATCAATGAAATTCT-GGGGGTGGGGGTGGGGCTGAAAGGGTG--GGGT 758
          |||||||||||||||||||||||||| |||||||||||||||| ||||||||||  || |
Db    630 GTTGGGAGAAATATCAATGAAATTCTAGGGGGTGGGGGTGGGGTTGAAAGGGTGCAGGTT 689

Qy    759 GGGAGATATGGAGGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCC 818
          |||||||||||||||||||||||||||||||||||||||  ||||   | |        |
Db    690 GGGAGATATGGAGGAGTATGAATTAGGGCTTGGAGTTGGGTAAAAACATTCCTGACTATC 749

Qy    819 TTCTTAACCACGTGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAAT 878
          |||||||||||||||||||||||||| |||||| ||||||| ||||||||||||||||
Db    750 TTCTTAACCACGTGGCTGATGTGGGGTAAGTATGAAGGGAAGGCAGTGGAGTAGCCTAAT 809

Qy    879 GAAAAGGGGTTCTAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACC 938
          |||||||| |     |    |||||||||||  ||  |||||||     | ||   |||
Db    810 GAAAAGGGTTCTAGTTGAGCCTCTGTAGATAATGGC--TGTTTCACGTGTGGTGAGAACC 867

Qy    939 TGGTGTCAGATAAAAGAAAC 958
          |||||||  ||||||  || |
Db    868 TGGTGTCGCATAAAACAATC 887
```

```
RESULT  15
BI753979
LOCUS       BI753979                765 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603027411F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5197803 5',
            mRNA sequence.
ACCESSION   BI753979
VERSION     BI753979.1  GI:15745557
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 765)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11495  row: e  column: 04
            High quality sequence stop: 722.
FEATURES             Location/Qualifiers
     source          1. .765
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5197803"
                     /clone_lib="NIH_MGC_114"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is  destroyed
                     upon cloning).  Average insert size 1.5 kb, insert size
                     range 1-3 kb.  Library is normalized and enriched for
                     full-length  clones and was constructed by C. Gruber
                     (Invitrogen).  Research  Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
BASE COUNT      207 a    152 c    228 g    178 t
ORIGIN
```

```
  Query Match             51.2%;  Score 613.4;  DB 10;  Length 765;
  Best Local Similarity   94.2%;  Pred. No. 2.7e-90;
  Matches  709;  Conservative    0;  Mismatches   21;  Indels   23;  Gaps    6;

Qy    454 GGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTTGCC 513
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db      1 GGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGA----------C 50

Qy    514 GCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCCTTC---------AG 564
          ||||||||||||||||||||||||||||||||||||||||||||||||||         ||
Db     51 GCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCA 110

Qy    565 CCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGT 624
          |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    111 GCCAGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGT 170

Qy    625 GTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTAC 684
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    171 GTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTAC 230

Qy    685 CTTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGTGGGGC 744
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 CTTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGTGGGGT 290

Qy    745 TGAAAGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAGGGCTTGGAGTTGGTAAAAAC 804
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 TGAAAGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAGGGCTTGGAGTTGGTAAAAAC 350

Qy    805 ATTCCTGACTATCCTTCTTAACCACGTGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAG 864
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 ATTCCTGACTATCCTTCTTAACCACGTGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAG 410

Qy    865 TGGAGTAGCCTAATGAAAAGGGGTTCTAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAG 924
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    411 TGGAGTAGCCTAATGAAAAGGGGTTCTAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAG 470

Qy    925 TGTGGTTGGAGACCTGGTGTCAGATAAAAGAAACTCCATCCGCACAGACAGATGCAAACA 984
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    471 TGTGGTTGGAGACCTGGTGTCAGATAAAAGAAACTCCATCCGCACAGACAGATGCAAACA 530

Qy    985 GCTCCTCTAGTTCTGCAGAGCTAGTTGAG-AACTCAACATTAATCATTTTAAAAAGTACT 1043
          | ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    531 G-TCCTCTAGTTCTGCAGAGCTAGTTGAGAAACTCAACATTAATCATTTTAAAAAGTACT 589

Qy   1044 GTCCTTGAAATAGATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTG 1103
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    590 GTCCTTGAAATAGATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTG 649

Qy   1104 AGCGTGGGGAACCCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAA 1163
          |||||  |||||||||||||||  |||||||||||||  |||||||||||||||||||  ||
Db    650 AGCGT-GGGAACCCCACAGAG-CCAAAGGACTTTTTAAGTATTCGAAATAAACAAAAGAA 707

Qy   1164 AAACCCATGAAAAAACCCAAAAAAAAAAAAAAA 1196
          |||||    |||||  |   |    | ||  |||| |
Db    708 AAACCATGAAAAAACCAAACGCAGAAGAAAACA 740


RESULT  16
BG742185
LOCUS       BG742185                703 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602634070F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4779400 5',
            mRNA sequence.
ACCESSION   BG742185
VERSION     BG742185.1  GI:14052838
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 703)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
            Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10636  row: k  column: 17
            High quality sequence stop: 701.
FEATURES             Location/Qualifiers
     source          1. .703
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4779400"
                     /clone_lib="NCI_CGAP_Skn3"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.5kb.  Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      171 a    213 c    189 g    130 t
ORIGIN


  Query Match             50.7%;  Score 608.4;  DB 10;  Length 703;
  Best Local Similarity   95.8%;  Pred. No. 1.8e-89;
  Matches  686;  Conservative    0;  Mismatches    6;  Indels   24;  Gaps    5;

Qy     10 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 69
          |||||||||||||||||| |||||||||||||||||||||||||||||||| ||||| |
Db      1 GCCGAAGCAGCAGCGCAGG-TGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGG-TGCCT 58

Qy     70 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     59 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 118

Qy    130 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    119 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 178

Qy    190 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 238

Qy    250 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    239 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 298

Qy    310 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 358

Qy    370 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    359 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 418

Qy    430 CTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTG 489
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    419 CTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTG 478

Qy    490 CCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGG 549
          |||||||||||||          ||||||||||||||||||||||||||||||||||| |
Db    479 CCAGATCCTGAGA----------CGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTTG 528

Qy    550 ----------CTCCTGCCCCTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGG 598
                     ||||||   | ||||||||||||||||||||||||||||||||||||||
Db    529 GCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGG 588

Qy    599 AGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACC 658
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    589 AGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACC 648

Qy    659 AAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATA 714
          ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    649 AAGACCTTGAACC-TTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATA 703


RESULT 17
BI195920
LOCUS       BI195920                927 bp    mRNA    linear   EST 10-JUN-2001
DEFINITION  602756190F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4891726 5',
            mRNA sequence.
ACCESSION   BI195920
VERSION     BI195920.1  GI:14650940
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 927)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1771  row: c  column: 23
            High quality sequence stop: 689.
FEATURES             Location/Qualifiers
     source          1. .927
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4891726"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      231 a    229 c    278 g    188 t      1 others
ORIGIN


  Query Match             50.5%;  Score 605;  DB 10;  Length 927;
  Best Local Similarity   87.0%;  Pred. No. 6e-89;
  Matches  734;  Conservative    0;  Mismatches  101;  Indels    9;  Gaps    6;

Qy      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      7 GTGACAGGAGCCGAAGCAGCAGGGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 66

Qy     61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     67 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 126

Qy    121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    127 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 186

Qy    181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    187 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 246

Qy    241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    247 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 306

Qy    301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    307 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 366

Qy    361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          |||||||||||||||||||||||||||||||||  |||| ||||||||||||  ||||
Db    367 AGGACCAGACAAGAACCTGGTGACTGGTGATCACANTCCCA-CCCACAGGATCTTGCCCA 425

Qy    421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    426 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCG 485

Qy    481 GGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTC 540
          |||   |||    |||||||| | |     | || |||  ||  | ||    | ||||
Db    486 GGGCTCTTGCAGATTCCTGAGACGGTTCCCCTCCCTGCCCACCCGGGTCCTGTGCTGGTC 545

Qy    541 CTGTGCTGGCTCCTGCCCCTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAG 600
          |||  |  |         ||   |||||||||||||||||||||||||||  |
Db    546 CTGGCC---CCTTCCTGTTTTGCAGCAGGGGTCAGGAGGTGGCTCGGGTGTGGGGCTGGA 602

Qy    601 AGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAA 660
            |   | ||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    603 GAGGCAGAAGCCTTTCCTGTTGGTGT-CCAGCACATGGAGCCCCTTGGGCTGAGCACCAA 661

Qy    661 GACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAATG 720
          ||| |  | | |||||   || || ||||||||||||||||||||||| ||
Db    662 GACTTGAACCTTTTTTGTGTTAACTTTTTTTCCAAATAACAGTTGGGACAATATCACTGA 721

Qy    721 AAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAG-ATATGGAGGAGTATGA 779
          || | ||||||| || ||||||| ||| | || |||||||||| |||||||||||| |||
Db    722 AATTCCTGGGGGGGGCGGTGGGGTTGACACGGGGGGGTGGGAGAATATGGAGGAGTTTGA 781

Qy    780 ATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACGTGGCTGATG 839
          |||||||  |||||||||||||||||||   | ||||  | ||||||||||  |||
Db    782 CTTAGGGC-TGGAGTTGGTAAAAACATTCTGAAGTATC--TTTTAACCACGTGCTTGAGT 838

Qy    840 TGGG 843
            |||
Db    839 GGGG 842


RESULT 18
BF981595
LOCUS       BF981595                1140 bp    mRNA    linear   EST 23-JAN-2001
DEFINITION  602309483F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4400637 5',
            mRNA sequence.
ACCESSION   BF981595
VERSION     BF981595.1  GI:12384407
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1140)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10106  row: e  column: 22
            High quality sequence stop: 598.
FEATURES             Location/Qualifiers
```

```
     source          1. .1140
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4400637"
                     /clone_lib="NIH_MGC_88"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      307 a    345 c    315 g    173 t
ORIGIN


  Query Match             49.3%;  Score 591.2;  DB 10;  Length 1140;
  Best Local Similarity   83.0%;  Pred. No. 9.7e-87;
  Matches  735;  Conservative    0;  Mismatches  143;  Indels    8;  Gaps    5;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           ||||||||||||||||||||||||||||||   |||  || |||  |||| | || |||||
Db       6 GTGACAGGAGCCGAAGCAGCAGCGCAGGTGTCCCCGTTTCCCCTCCCCCTTCCCGTTCTC 65

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 125

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     126 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 185

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     186 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 245

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     246 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 305

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     306 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 365

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     366 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 425

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     426 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 485

Qy     481 GGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCC-TCCCTGCCCCACCCGGGT 539
           ||||||||||||||||||||    |    | || ||||  ||| |  ||      |
Db     486 GGGGCTCTGCCAGATCCTGAGAAGCTTCCCCTCCCTGCCCCAGCCCGGGTCCTGTGCTGG 545

Qy     540 CCTGTGCTGGCTCCTGCCCCTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGG-CTGG 598
           |   |||    ||||||   | ||||||||||||||||||||||||||||||||| ||||
Db     546 CTACTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCCTGG 605

Qy     599 AGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACC 658
           | |||||||||||||||||||||||| ||  |   ||||||||||  || ||||||||
Db     606 ACAGGCAGAAGCCCTTTCCTGTTGGTGCCCATG--AATGGAGCCCCTGAGGTTGAGCACC 663

Qy     659 AAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAA 718
           |||||  ||||||   ||  | || ||| ||||| | ||||||   ||||| |||  || |
Db     664 AAGACTTTGAAC---TTACGGTTAACCCTTTTTTCCAATAACCAGTGGGA-AAACTTCCA 719

Qy     719 TGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGGAGGAGTATG 778
           |||||    |||    | |  |||    | |||| |||||||     |||| |  ||
Db     720 TGAAACCTGGGGAAGAGAGACGGGTTGAACAGGGAGGGGTGGACACAATGGGGCGGTTGA 779

Qy     779 AATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACGTGGCTGAT 838
           |  | ||| | ||  ||| |||| |||||| || ||   ||    ||   |  ||  |
Db     780 CAATGAGGCATTGAACTGGGAAAACAATTCCGGAATACTTTTAAACACACAGCAGCATAC 839

Qy     839 GTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAG 884
           | ||| |||| || | | | |      |  ||||||  |  |  ||
Db     840 GGGGGCTAGGCATCACGCGGAAAGCACGAGGTAGCCGCACCAGCAG 885


RESULT  19
AK006149
LOCUS       AK006149                1139 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus adult male testis cDNA, RIKEN full-length enriched
            library, clone:1700020C18:endosulfine alpha, full insert sequence.
ACCESSION   AK006149
VERSION     AK006149.1  GI:12839095
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:1700020C18.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1139)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
```

```
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence[5'
            GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA
            was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3'
            end: SstI. Host: SOLR.
FEATURES              Location/Qualifiers
     source           1..  .1139
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="MGD:MGI:1908876"
                      /db_xref="taxon:10090"
                      /clone="1700020C18"
                      /sex="male"
                      /tissue_type="testis"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     gene             111. .476
                      /gene="Ensa"
     CDS              111. .476
                      /gene="Ensa"
                      /note="data source:MGD, source key:MGI:1891189,
                      evidence:ISS
                      endosulfine alpha
                      putative"
                      /codon_start=1
                      /protein_id="BAB24433.1"
                      /db_xref="GI:12839096"
                      /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPEKAEEAKLKAKYPSL
                      GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGADKNLVTGDHIPTPQD
                      LPQRKSSLVTSKLAGGQVE"
BASE COUNT      310 a    267 c    314 g    248 t
ORIGIN


  Query Match           49.1%;  Score 589.2;  DB 11;  Length 1139;
  Best Local Similarity  77.3%;  Pred. No. 2e-86;
  Matches  850;  Conservative    0;  Mismatches  198;  Indels   52;  Gaps    9;

Qy      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          |||||||||||||| |||||||| |||||||||||||||||||||||||||||||||| ||
Db      7 GTGACAGGAGCCGGAGCAGCAGCTCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTTTC 66

Qy     61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||| || ||||||||||||  | || | |||||||||||| | ||||||||||||||||||
Db     67 CGGCTGACTTCCCGGGCCC--TGCATTACACAGTCCCGGTTCTGCCATGTCCCAGAAACA 124

Qy    121 AGAAGAAGAGAAGCCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||| ||||||||||||||||||||||||||||||||| || ||||||||||||
Db    125 AGAAGAAGAAAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGATACACAGGAGAAAGA 184

Qy    181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||| ||||| ||||||| |||||| |||||||||||||||||||| ||||||||||
Db    185 AGGGATTCTCCCTGAGAAAGCTGAGGAGGCAAAGCTAAAGGCCAAATATCCAAGCCTAGG 244

Qy    241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    245 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 304

Qy    301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||

Db    305 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAACAAGCAGCTGCCAAGTGC 364

Qy    361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          |||| |||||||||||||||||| ||||| |||||||||||||||||| ||||||||
Db    365 AGGAGCAGACAAGAACCTGGTGACCGGTGACCACATCCCCACCCCACAGGACCTGCCCCA 424

Qy    421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    425 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAGTGATGCTTCCT 484

Qy    481 GGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTC 540
          ||||||| ||||||| ||||||              | ||| | | |  ||      | |
Db    485 GGGGCTCCACCAGATCGTGAGAC-----------GCTGCCCACTCGGGTCCTGTGCTGGC 533

Qy    541 CTGTGCTGGCTCCTGCCCCTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAG 600
             |||     |||||   |  |||||||||||||||||||| |||||| |||| |||
Db    534 TCCTGCCCCTCCCTGCTTTTGTAGCCAGGGGTCAGGAGGTGGCCCGGGTGCAGGCTAGAG 593

Qy    601 AGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAA 660
          |||||||   |||  ||   |||||||||| ||||||||||||||||||||||||||||
Db    594 AGGCAGAGAACCTGCCCAC-TGGTGTCCCAACACATGGAGCCCCTTGGGCTGAGCACCAA 652

Qy    661 GACCTTGAACCTTTTTT-------GTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAAT 713
          |||||||||| ||||||        |||  | |||||||||||||| ||||| |||||
Db    653 GACCTTGAACTTTTTTTTGTTTTGTTTTACCTTTTTTTCCAAATAACTGTTGGAAGAAAA 712

Qy    714 ATCAATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGGAGGA 773
           ||| ||| || |         |||||||| ||||||| ||| |||    | |||  ||
Db    713 CTCATTGAGATCC------CAAGGGTGGGGTTGAAAGGATGGAGTG---AAGATGTGGGG 763

Qy    774 GTATGAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACGTGG 833
          |  ||| | |||| ||     || ||||||      |  ||| | ||||     || |
Db    764 GAGTGAGTGAGGGTTTATGACTGATAAAAAAGAAGTTTCCTACCGTTCTCCTGAACCTCT 823

Qy    834 CTGATGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTAG 893
            | ||  | |   |  ||   ||   |               ||| |||||||||  |
Db    824 GGGTTGCTGTGGTTGACTGTAAGGTGAGGG------------ATGGTAAGGGGTTCAGG 870

Qy    894 TTGAGCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATAAAA 953
          || ||||||| ||| | | ||  | ||| | | ||| | | || || |||||||| ||||
Db    871 TTCAGCTCTGAAGAGAGAAGCACTCTTTGAATATGGCTAGTGATCTAGTGTCAGAGAAAA 930

Qy    954 GAAACTCCATCCGCACAGACAGATGCAAACAGCTCCTCTAGTTCTGCAGAGCTAGTTGAG 1013
          | || |   | |    |  |||| |      |   |||| |||||||||| | |||||||
Db    931 GGAATTGTGTTCTTTAAAGCAGACG----AGGTCTCTCTGGTTCTGCAGAACAAGTTGAG 986

Qy   1014 AACTCAACATTAAT-----CATTTTAAAAAGTACTGTCCTTGAAATAGATTTGCTGTGGG 1068
          | | |  || ||      | ||| ||||| | |||||||||||||| ||||||||| |
Db    987 GATTTAGGATAAACTGTTCCTTTTAAAAAAATGCTGTCCTTGAAATAGGTTTGCTGTGTG 1046

Qy   1069 AAGAAGGGCAGTGAGTGTGG 1088
           |||||||||||||| || ||
Db   1047 GAGAAGGGCAGTGAATGGGG 1066


RESULT  20
BI222786
LOCUS       BI222786                824 bp    mRNA    linear   EST 11-JUL-2001
DEFINITION  602941568F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5104578 5',
            mRNA sequence.
ACCESSION   BI222786
VERSION     BI222786.1  GI:14676230
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 824)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
          Tissue Procurement: ATCC
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: Incyte Genomics, Inc.
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM11252  row: h  column: 18
          High quality sequence stop: 688.
FEATURES             Location/Qualifiers
     source          1. .824
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5104577"
                     /clone_lib="NIH_MGC_12"
                     /tissue_type="cervical carcinoma cell line"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.4 kb. Library prepared by Life
                     Technologies."
BASE COUNT      188 a    256 c    226 g    154 t
ORIGIN
```

```
  Query Match             48.8%;  Score 585.6;  DB 10;  Length 824;
  Best Local Similarity   88.8%;  Pred. No. 8.7e-86;
  Matches  691;  Conservative    0;  Mismatches   79;  Indels    8;  Gaps    5;

Qy       6 AGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCC-TCCCCCTTCCCTTCTCCGGT 64
           |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db       1 AGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCGTCCCCCTTCCCTTCTCCGGT 60

Qy      65 TGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAA 124
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAA 120

Qy     125 GAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAA-GAAGG 183
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     121 GAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAACGAAGG 180

Qy     184 TATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACA 243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACA 240

Qy     244 AAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGA 303
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGA 300

Qy     304 CTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGG 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGG 360

Qy     364 ACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAG 423
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     361 ACCAGACAAGAACCTGGTGACTGGTGATCACATTCCCACCCCACAGGATCTGCCCCAGAG 420

Qy     424 AAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGG 483
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGG 480

Qy     484 GCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTG 543
           |||||||||||||||||| ||||   |  | | |||  || |  ||       | |
Db     481 GCTCTGCCAGATCCTGAGAC-GCTTCCCCTCCCTGCCCCACCGGGTCCTGTGCTGGCTCC 539

Qy     544 TGCTGGCTCCTGCCCCTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGG 603
           |||    ||||||    | |||||||| ||||||||||||||||||||||||||||| |||
Db     540 TGCCCCTTCCTGCTTTTGCAGCCAGGCGTCAGGAGGTGGCTCGGGTGTGGGCTGGACAGG 599

Qy     604 CAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGAC 663
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |   |
Db     600 CAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACAAGACC 659
```

```
Qy     664 CTTGAACCTTTTTT--GTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATC---AA 718
             |  | |||| | |   |||  ||||||||||||||||||| | ||  | | |||
Db     660 TTGAAGCCTTCTCTCGTTTTACCCTTTTTTCCAAATAACAGCTCGGGCACACATCTCACT 719

Qy     719 TGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGGAGGAGTA 776
             | | |||||||||| |||||||  |  |   |   || | | || |    |||| |
Db     720 GACACTACTGGGGGTGAGGGTGGGCGTTGACAAGGCTGGGGTGCGACACTCTGGAGGA 777
```

```
RESULT  21
BE906167
LOCUS       BE906167                606 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601497230F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3899287 5',
            mRNA sequence.
ACCESSION   BE906167
VERSION     BE906167.1  GI:10399509
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 606)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9696  row: l  column: 08
            High quality sequence stop: 606.
FEATURES             Location/Qualifiers
     source          1. .606
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3899287"
                     /clone_lib="NIH_MGC_70"
                     /tissue_type="epithelioid carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.1 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      175 a    101 c    179 g    151 t
ORIGIN
```

```
  Query Match             48.4%;  Score 580.8;  DB 10;  Length 606;
  Best Local Similarity   98.7%;  Pred. No. 5.8e-85;
  Matches  596;  Conservative    0;  Mismatches    7;  Indels    1;  Gaps    1;

Qy     577 AGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACAT 636
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACAT 60

Qy     637 GGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAA 696
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAA 120

Qy     697 TAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGG 756
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     121 TAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGTGGGGTTGAAAGGGTGGG 180

Qy     757 GTGGGAGATATGGAGGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTAT 816
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GTGGGAGATATGGAGGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTAT 240

Qy     817 CCTTCTTAACCACGTGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTA 876
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCTTCTTAACCACGTGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTA 300

Qy    877 ATGAAAAGGGGTTCTAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGA 936
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ATGAAAAGGGGTTCTAGTTGAGCTCTGTAGATAAATGCCTTGTTTCAGTGTGGTTGGAGA 360

Qy    937 CCTGGTGTCAGATAAAAGAAACTCCATCCGCACAGACAGATGCAAACAGCTCCTCTAG-T 995
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    361 CCTGGTGTCAGATAAAAGAAACTCCATCCGCACAGACAGATGCAAACAGCTCCTCTAGTT 420

Qy    996 TCTGCAGAGCTAGTTGAGAACTCAACATTAATCATTTTAAAAAGTACTGTCCTTGAAATA 1055
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCTGCAGAGCTAGTTGAGAACTCAACATTAATCATTTTAAAAAGTACTGTCCTTGAAATA 480

Qy   1056 GATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAAC 1115
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GATTTGCTGTGGGAAGAAGGGCAGTGAGTGTGGGAGAAAGGAGCCGTGAGCGTGGGGAAC 540

Qy   1116 CCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAACAAAAACCCATGAAA 1175
          ||||||||||||||||||||||||||||||||||||||||| ||||||||    |||
Db    541 CCCACAGAGCCCAAAGGACTTTTTCAGTATTCGAAATAAACAAAAAAAAAACCATGAAAA 600

Qy   1176 AAAC 1179
          || |
Db    601 AACC 604


RESULT  22
BG741487
LOCUS       BG741487                 688 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602632285F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4777461 5',
            mRNA sequence.
ACCESSION   BG741487
VERSION     BG741487.1  GI:14052140
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 688)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
            Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10631  row: j  column: 22
            High quality sequence start: 5
            High quality sequence stop: 686.
FEATURES             Location/Qualifiers
     source          1. .688
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4777461"
                     /clone_lib="NCI_CGAP_Skn3"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.5kb.  Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      164 a    212 c    184 g    128 t
ORIGIN


  Query Match             48.2%;  Score 577.8;  DB 10;  Length 688;
  Best Local Similarity   94.4%;  Pred. No. 1.7e-84;
```

```
  Matches  654;  Conservative    0;  Mismatches   32;  Indels    7;  Gaps    5;

Qy     10 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 69
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 60

Qy     70 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 120

Qy    130 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 189
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    121 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTA-TCT 179

Qy    190 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 249
          |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    180 GCCTGAGAGAGCTGAA-AGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 238

Qy    250 TGGAGGCTCCGACTTCCTCATGAAGAGACTCC--AGAAAGGGCAAAAGTACTTTGACTCA 307
          |||||||||||||||||||||||||||||||    |||||||||||||||||||||||||
Db    239 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCCAGGAAAGGGCAAAAGTACTTTGACTCA 298

Qy    308 GGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCA 367
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299 GGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCA 358

Qy    368 GACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAG 427
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    359 GACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAG 418

Qy    428 TCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTC 487
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    419 TCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTC 478

Qy    488 TGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCT 547
          ||||||||||||||||    |      | || |||  |    ||    | || |||  |||
Db    479 TGCCAGATCCTGAGACGCTTCCCCTCCCTGCCCACCGGGTCCTGTGCTGGCTCC--TGCC 536

Qy    548 GGCTCCTGCCCCTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGA 607
             ||||||   | ||||||||||||||||||||||||||||||||||||||||||||||||
Db    537 CCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGA 596

Qy    608 AGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTG 667
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    597 AGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCATGACCTTG 656

Qy    668 AACCTTTTTTGTTTTACCTTTTTTCCAAATAAC 700
          |||| |||||||||||||||||||||||||||
Db    657 AACC-TTTTTGTTTTACCTTTTTTCCAAATAAC 688


RESULT  23
BF795934
LOCUS       BF795934                 972 bp    mRNA    linear   EST 12-JAN-2001
DEFINITION  602259188F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4342533 5',
            mRNA sequence.
ACCESSION   BF795934
VERSION     BF795934.1  GI:12100988
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 972)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis Staudt, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
```

```
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9957  row: p  column: 22
            High quality sequence stop: 639.
FEATURES             Location/Qualifiers
     source          1. .972
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4342533"
                     /clone_lib="NIH_MGC_85"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.867 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      232 a    288 c    273 g    179 t
ORIGIN


  Query Match             48.2%;  Score 577.6;  DB 10;  Length 972;
  Best Local Similarity   95.5%;  Pred. No. 1.6e-84;
  Matches  644;  Conservative    0;  Mismatches    9;  Indels   21;  Gaps    4;

Qy      28 GTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTT-GCCTTCCCGGGCCCCTTACAC 86
           ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db       1 GTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGGCCTTCCCGGGCCCCTTACAC 60

Qy      87 TCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGA 146
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGA 120

Qy     147 CCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAG 206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAG 180

Qy     207 AGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCC 266
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCC 240

Qy     267 TCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCA 326
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCA 300

Qy     327 AAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTG 386
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTG 360

Qy     387 GTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCA 446
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCA 420

Qy     447 AGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGC 506
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGA---- 476

Qy     507 TTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCCTTC---- 562
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     477 ------CGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTCCTGC 530

Qy     563 -----AGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCC 617
                  |  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     531 TTTTGCAGCCAGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCC 590

Qy     618 TGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTT 677
           |||||||||||||||||||||| |||||||||||||||||||||||||| |||| ||
Db     591 TGTTGGTGTCCCAGCACATGGAG-CCCTTGGGCTGAGCACCAAGACCTTGAAGCTTTGTT 649

Qy     678 GTTTTACCTTTTTT 691
           ||||| || |||||
Db     650 GTTTTTCCCTTTTT 663
```

```
RESULT  24
BM478401
LOCUS       BM478401                1097 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6507768 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5578513
            5', mRNA sequence.
ACCESSION   BM478401
VERSION     BM478401.1  GI:18527443
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1097)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12333  row: l  column: 02
            High quality sequence stop: 550.
FEATURES             Location/Qualifiers
     source          1. .1097
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5578513"
                     /clone_lib="NIH_MGC_67"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      199 a    385 c    295 g    218 t
ORIGIN


  Query Match             48.2%;  Score 577.4;  DB 10;  Length 1097;
  Best Local Similarity   92.5%;  Pred. No. 1.7e-84;
  Matches  658;  Conservative    0;  Mismatches   31;  Indels   22;  Gaps    4;

Qy      16 GCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGG 75
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13 GCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGG 72

Qy      76 GCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCC 135
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73 GCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCC 132

Qy     136 TGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGA 195
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     133 TGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGA 192

Qy     196 GAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGG 255
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     193 GAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGG 252

Qy     256 CTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTA 315
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     253 CTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTA 312

Qy     316 CAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAA 375
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     313 CAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAA 372
```

```
Qy    376 CCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCT 435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    373 CCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCT 432

Qy    436 CGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGAT 495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    433 CGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGAT 492

Qy    496 CCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTG 555
          |||||||          ||||||||||||||||||||||||||||||||||||||||||
Db    493 CCTGAGA----------CGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTG 542

Qy    556 CCCCTT----------CAGCCAGGGGTCAGGAGGT-GGCTCGGGTGTGGGCTGGAGAGGC 604
          ||||||          |||||||||| ||||| || ||||||||||| |||||| ||||
Db    543 CCCCTTCCTGCTTTTGCAGCCAGGGGCCAGGAAGTGGGCTCGGGTGTGTGCTGGACAGGC 602

Qy    605 AGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACC 664
          |||| ||||||||||| |||| ||| |||   ||||||||||||||||||||||| ||| |
Db    603 AGAACCCCTTTCCTGTGGGTGCCCCCGCAATGGGAGCCCCTTGGGCTGAGCACCCAGAGC 662

Qy    665 TTGAA-CCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATA 714
          ||||| || |||| |  ||| ||   ||||| |  ||  || ||  | | |
Db    663 TTGAACCCCTTTTCGACTTAGCTGGCTTCCATACCACCCTTTGGGTATACA 713


RESULT 25
BG716405
LOCUS       BG716405                731 bp    mRNA    linear   EST 08-MAY-2001
DEFINITION  602676303F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4798825 5',
            mRNA sequence.
ACCESSION   BG716405
VERSION     BG716405.1  GI:13995592
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 731)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10687  row: e  column: 02
            High quality sequence stop: 727.
FEATURES             Location/Qualifiers
     source          1. .731
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4798825"
                     /clone_lib="NIH_MGC_96"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.3 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length  clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      185 a    212 c    197 g    137 t
ORIGIN
```

```
  Query Match           48.0%;  Score 575.8;  DB 10;  Length 731;
  Best Local Similarity  93.4%;  Pred. No. 3.5e-84;
  Matches  668;  Conservative    0;  Mismatches   37;  Indels   10;  Gaps    6;

Qy      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     26 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 85

Qy     61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||| |||||
Db     86 CGGTTGCCTTCCCGGG-CCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCA-AAACA 143

Qy    121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    144 AGAAGAAGAAAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 203

Qy    181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    204 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 263

Qy    241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    264 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 323

Qy    301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    324 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 383

Qy    361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    384 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 443

Qy    421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    444 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGGCC 503

Qy    481 GGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTC 540
          ||||||||||||||||||||  ||||   |  | | ||| ||| |  ||      | |
Db    504 GGGGCTCTGCCAGATCCTGAGA--GCTTCCCCTCCCTGCCCACCCGGGTCCTGTGCTGGC 561

Qy    541 CTGTGCTGGCTCCTGCCCCTTCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAG 600
            |||   ||||||   | ||||||||||||||||||||||||| ||||||||||||
Db    562 TCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGATGTGGGCTGGAG 621

Qy    601 AGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAA 660
          ||||||||||||||||||||||||||||   |||||||||||||||||||||||||| |
Db    622 AGGCAGAAGCCCTTTCCTGTTGGTGTCC--ACACATGGAGCCCCTTGGGCTGAGCACAAG 679

Qy    661 GACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATAT 715
            |  ||||||||||||   ||||||||||||||||||||||||||||||||||
Db    680 AA--TTGAACCTTTTT--GTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATAT 730


RESULT 26
BI835836
LOCUS       BI835836                670 bp    mRNA    linear   EST 04-OCT-2001
DEFINITION  603087830F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5226774 5',
            mRNA sequence.
ACCESSION   BI835836
VERSION     BI835836.1  GI:15947386
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 670)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
```

```
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM11570  row: l  column: 07
           High quality sequence stop: 619.
FEATURES             Location/Qualifiers
     source          1. .670
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5226774"
                     /clone_lib="NIH_MGC_120"
                     /lab_host="DH10B"
                     /note="Organ: pooled pancreas and spleen; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of spleen and pancreas from 28 yo
                     male.  Library is oligo-dT primed and directionally cloned
                     (EcoRV  site is destroyed upon cloning).  Average insert
                     size 1.5 kb, insert size range 1-2.5 kb.  Library is
                     normalized and enriched for  full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 025. Note: this is a NIH_MGC Library."
BASE COUNT      160 a    204 c    181 g    125 t
ORIGIN


  Query Match             47.9%;  Score 574.8;  DB 10;  Length 670;
  Best Local Similarity   94.1%;  Pred. No. 5.3e-84;
  Matches  638;  Conservative    0;  Mismatches   17;  Indels   23;  Gaps    3;

Qy      18 AGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGC 77
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGC 60

Qy      78 CCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTG 137
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTG 120

Qy     138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 180

Qy     198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 240

Qy     258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 300

Qy     318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 360

Qy     378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 420

Qy     438 TCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCC 497
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCC 480

Qy     498 TGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCC 557
           |||||          |||||||||||||| ||||||||||||||||||||||||||||
Db     481 TGAGA----------CGCTTCCCCTCCCTGTCCCACCCGGGTCCTGTGCTGGCTCCTGCC 530

Qy     558 CCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGA 607
           ||||          ||| |||| ||||||||||||||||||||||||||||||||||||
Db     531 CCTTCCTGCTTTTGCAGTCAGGCGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGA 590

Qy     608 AGCCCTTTCCTGTTGGTGTCCCAGCACATGGAG---CCCCTTGGGCTGAGCACCAAGACC 664
           |||||||||||||||||||||||||||||||     |||   ||| | ||    |||||||||| |
Db     591 AGCCCTTTCCTGTTGGTGTCCCAGCTACATGAGAGCCCCTTGGGCTGAAGCACCAAGAGC 650

Qy     665 TTGAACCTTTTTTGTTTT 682
           ||||||||||||  |||
Db     651 TTGAACCTTTTTTCGTTT 668


RESULT  27
BI223805
LOCUS       BI223805                713 bp    mRNA    linear   EST 11-JUL-2001
DEFINITION  602942709F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5105571 5',
            mRNA sequence.
ACCESSION   BI223805
VERSION     BI223805.1  GI:14677249
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 713)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Incyte Genomics, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM11255  row: b  column: 04
           High quality sequence stop: 648.
FEATURES             Location/Qualifiers
     source          1. .713
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5105571"
                     /clone_lib="NIH_MGC_12"
                     /tissue_type="cervical carcinoma cell line"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.4 kb. Library prepared by Life
                     Technologies."
BASE COUNT      189 a    206 c    188 g    130 t
ORIGIN


  Query Match             47.9%;  Score 574.4;  DB 10;  Length 713;
  Best Local Similarity   90.6%;  Pred. No. 6e-84;
  Matches  647;  Conservative    0;  Mismatches   61;  Indels    6;  Gaps    3;

Qy      27 GGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCCCTTACAC 86
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCCCTTACAC 60

Qy      87 TCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGA 146
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGA 120

Qy     147 CCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAG 206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAG 180

Qy     207 AGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCC 266
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCC 240

Qy     267 TCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCA 326
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    241 TCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCA 300

Qy    327 AAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTG 386
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTG 360

Qy    387 GTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCA 446
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCA 420

Qy    447 AGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGC 506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACGCT 480

Qy    507 TTTTGCCGCTTCCCCTCCC----TGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTC 562
          |    | || |||| |||    | |    | || |||||  |    |   |  | |
Db    481 TCCCCTCCCTGCCCCACCCGGGATCCTGTGACTGGCTCCTGCACCCTTCCTGACTTTTGC 540

Qy    563 AGCCAGGGGTCAGGAGGTGG-CTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTT 621
          || ||||||||||||||||| ||||||||||||||||||||| ||||||||||   |
Db    541 AGACAGGGGTCAGGAGGTGGACTCGGGTGTGGGCTGGAGAGGCAAAAGCCCTTTCCATGT 600

Qy    622 GGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTT 681
          |||||||    |||||||| |||||||||||||   |||||||||||||| ||   |||
Db    601 GGTGTCCAGCAACATGGAGACCCTTGGGCTGAGACACAAGACCTTGAACC-TTAAAGTTA 659

Qy    682 TACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGG 735
          ||| | ||| | | ||||||||||||| |||||||||||||||    | |||||
Db    660 TACTTATTTCCAACATAACAGTTGGGACAAATATCAATGAAATCTGAGAGGTGG 713


RESULT  28
BG180105
LOCUS       BG180105                975 bp    mRNA    linear   EST 06-FEB-2001
DEFINITION  602329682F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4430925 5',
            mRNA sequence.
ACCESSION   BG180105
VERSION     BG180105.1  GI:12686808
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 975)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10185  row: c  column: 22
            High quality sequence stop: 601.
FEATURES             Location/Qualifiers
     source          1. .975
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4430925"
                     /clone_lib="NIH_MGC_91"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.4 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      275 a    187 c    270 g    243 t
ORIGIN
```

```
  Query Match             47.7%;  Score 571.6;  DB 10;  Length 975;
  Best Local Similarity   89.2%;  Pred. No. 1.5e-83;
  Matches  692;  Conservative    0;  Mismatches   59;  Indels   25;  Gaps    6;

Qy    208 GGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAA-AGCCTGGAGGCTCCGACTTCC 266
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db      1 GGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAGAGCCTGGAGGCTCCGACTTCC 60

Qy    267 TCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCA 326
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCA 120

Qy    327 AAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTG 386
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTG 180

Qy    387 GTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCA 446
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCA 240

Qy    447 AGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGC 506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGA---- 296

Qy    507 TTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTC---- 562
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    297 ------CGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTCCTGC 350

Qy    563 -----AGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCC 617
                  | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 TTTTGCAGCCAGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCC 410

Qy    618 TGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTT 677
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    411 TGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTT 470

Qy    678 GTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGG 737
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    471 GTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGG 530

Qy    738 GTGGGGCTGAA-AGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAGGGCTTGGAGTTG 796
          |||||| |||| |||||||||||||||||||||||||||||||||||||||||| |   | |
Db    531 GTGGGGTTGAACAGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAGGGCTGGAGTTGG 590

Qy    797 GTAAAAACATTCCTGACTATCCTTCTTAACCACGTGGCTGATGTGGGGTAGGTATGAGGG 856
          || ||||   | |    ||| |||||||||     |       ||||||| |||||||||
Db    591 GTTAAAAACATTCCTGATATTCTTCTTAACACGTGGCTGATGTTGGGGTAAGTATGAGGG 650

Qy    857 GAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTA--GTTGAGCTCTGTAGATAAATGC 914
          ||||||| ||||||   |||||||||||||||| ||   |  | |||||||||| | |||
Db    651 GAAGGAATTGGAGTTAGCTAATGAAAAGGGGTTTTAAGTTGAAACTCTGTAGATTACTGC 710

Qy    915 CTTGTTTCAGTGTGGTTGGAGACCTGGTGTCAGATAAAAGAAACTCCATCCGCACA 970
          |||||||| |||||  | | || ||||||   ||| ||||||  ||  || ||| |
Db    711 CTTGTTTCGGTGTG--TTGGGAACTGGTGGTCGATTAAAGAACTTCATTCGGCAGA 764


RESULT  29
AU137227
LOCUS       AU137227                808 bp    mRNA    linear   EST 24-OCT-2000
DEFINITION  AU137227 PLACE1 Homo sapiens cDNA clone PLACE1006040 5', mRNA
            sequence.
ACCESSION   AU137227
VERSION     AU137227.1  GI:10997766
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 808)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
```

```
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .808
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PLACE1006040"
                     /clone_lib="PLACE1"
                     /tissue_type="placenta"
                     /note="Vector: pME18SFL3"
BASE COUNT      198 a    218 c    224 g    159 t      9 others
ORIGIN
```

```
  Query Match             47.6%;  Score 570.8;  DB 9;  Length 808;
  Best Local Similarity   94.1%;  Pred. No. 2.2e-83;
  Matches  664;  Conservative    0;  Mismatches   16;  Indels   26;  Gaps    6;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 80

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 140

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 200

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 260

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 320

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 380

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 440

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     441 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCC 500

Qy     481 GGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCC-ACCCGGGT 539
           ||||||||||||||||||||||          ||||||||||||||| | ||| ||||
Db     501 GGGGCTCTGCCAGATCCTGAGA----------CGCTTCCCCTCCCTGCCNCAACCGGGGT 550

Qy     540 CCTGTGCTGGCTCCTGCCCCTTC----------AGCCAGGGGTCAGGAGGTGGCTCGGG 588
           ||||||||||||||||||||||          ||||||||||||||||||||||||||
Db     551 CCTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAAGCCAGGGGTCAGGAGGTGGCTCGGG 610

Qy     589 TGT--GGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTT 646
           |||  ||   |   ||||| || |||||||||||||||||| ||| ||||||||||||||
```

```
Db     611 TGTTGGGCTGGAGAAGGCANAANCCCTTTCCTGTTGGTGT-CCANCACATGGAGCCCCTT 669

Qy     647 GGGCTGAGCACC-AAGACCTTGAACCTTTTTTGTTTTACCTTTTTT 691
           |||||||||||| |||||||||| | ||||||||||  | |||||
Db     670 GGGCTGAGCACCAAAGACCTTGAAACCTTTTTGTTTTTACCTTTTT 715


RESULT  30
BE882748
LOCUS       BE882748                 647 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601505521F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3907277 5',
            mRNA sequence.
ACCESSION   BE882748
VERSION     BE882748.1  GI:10331524
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 647)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9717  row: i  column: 06
            High quality sequence stop: 643.
FEATURES             Location/Qualifiers
     source          1. .647
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3907277"
                     /clone_lib="NIH_MGC_71"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2.1 kb.  "
BASE COUNT      164 a    187 c    176 g    120 t
ORIGIN
```

```
  Query Match             47.2%;  Score 566.4;  DB 10;  Length 647;
  Best Local Similarity   95.9%;  Pred. No. 1.2e-82;
  Matches  631;  Conservative    0;  Mismatches    6;  Indels   21;  Gaps    4;

Qy      65 TGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAA 124
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAA 60

Qy     125 GAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGT 184
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGT 120

Qy     185 ATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAA 244
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAA 180

Qy     245 AAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGAC 304
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGAC 240

Qy     305 TCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGA 364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGA 300
```

```
Qy    365 CCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGA 424
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||.
Db    301 CCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGA 360

Qy    425 AAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGG 484
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGG 420

Qy    485 CTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGT 544
          |||||||||||||||||          |||||||||||||| ||||||||||||||||
Db    421 CTCTGCCAGATCCTGAGA----------CGCTTCCCCTCCCTG-CCCACCCGGGTCCTGT 469

Qy    545 GCTGGCTCCTGCCCCTTC---------AGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGC 595
          |||||||||||||||||            |  ||||||||||||||||||||||||||||
Db    470 GCTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGTCAGGAGGTGGCTCGGGTGTGGGC 529

Qy    596 TGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGC 655
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    530 TGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCTGAGC 589

Qy    656 ACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAAT 713
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||| |
Db    590 ACCAAGA-CTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAATT 646


RESULT  31
BI521459
LOCUS       BI521459                719 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  603081263F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5220663 5',
            mRNA sequence.
ACCESSION   BI521459
VERSION     BI521459.1  GI:15346251
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 719)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11554  row: m  column: 16
            High quality sequence stop: 716.
FEATURES             Location/Qualifiers
     source          1. .719
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5220663"
                     /clone_lib="NIH_MGC_120"
                     /lab_host="DH10B"
                     /note="Organ: pooled pancreas and spleen; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of spleen and pancreas from 28 yo
                     male.  Library is oligo-dT primed and directionally cloned
                     (EcoRV  site is destroyed upon cloning).  Average insert
                     size 1.5 kb, insert size range 1-2.5 kb.  Library is
                     normalized and enriched for  full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 025. Note: this is a NIH_MGC Library."
BASE COUNT      183 a    212 c    198 g    126 t
ORIGIN


  Query Match             47.2%;  Score 566;  DB 10;  Length 719;
  Best Local Similarity   93.4%;  Pred. No. 1.4e-82;
  Matches  676;  Conservative    0;  Mismatches   20;  Indels   28;  Gaps    7;

Qy      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     11 GTGACAGGAGCCGAAGCAGCAGCGCAGG-TGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 69

Qy     61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 129

Qy    121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    130 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 189

Qy    181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    190 AGGTATTCTGCCTGAGAGAGCTGAACAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 249

Qy    241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    250 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 309

Qy    301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 369

Qy    361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCA-CAGGATCTGCCCC 419
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    370 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCAGCAGGATCTGCCCC 429

Qy    420 AGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCC 479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    430 AGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCC 489

Qy    480 CGGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGT 539
          ||||||||||||||||||||||          |||||||||||||| |||| ||||||
Db    490 CGGGGCTCTGCCAGATCCTGAGA----------CGCTTCCCCTCCCTGACCCAACCGGGT 539

Qy    540 CCTGTGCTGG-----------CTCCTGCCCCTTCAGCCAGGGGTCAGGAGGTGGCTCGGG 588
          ||||||||||            ||||||   | ||||||||||||||||||||||||||||
Db    540 CCTGTGCTGGGGTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGG 599

Qy    589 TGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGG 648
          ||||||||||||||||||  |||||||||| ||||| |||||||||||||| |||||||
Db    600 TGTGGGCTGGAGAGGCAGAAAGCCTTTCCTGTAGGTGT-CCAGCACATGGAG-CCCTTGG 657

Qy    649 GCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGA 708
          ||||||||||||||||||||||   |||   |   | ||||| | ||||||||||||||
Db    658 GCTGAGCACCAAGACCTTGAACC---TTTAGGTATACCTTTTTTCCAATAACAGTTGGGA 714

Qy    709 GAAA 712
          ||||
Db    715 GAAA 718


RESULT  32
AV727450
LOCUS       AV727450                729 bp    mRNA    linear   EST 17-OCT-2000
DEFINITION  AV727450 HTC Homo sapiens cDNA clone HTCBGH11 5', mRNA sequence.
ACCESSION   AV727450
VERSION     AV727450.1  GI:10836871
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 729)
  AUTHORS   Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
            Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
            ,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
            Chen,J., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA HTC clones
```

```
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .729
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HTCBGH11"
                     /clone_lib="HTC"
                     /tissue_type="Hypothalamus"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      190 a    193 c    185 g    157 t      4 others
ORIGIN

  Query Match             46.8%;  Score 560.6;  DB 9;  Length 729;
  Best Local Similarity   95.0%;  Pred. No. 1e-81;
  Matches  606;  Conservative    0;  Mismatches   12;  Indels   20;  Gaps    2;

Qy    114 AGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGG 173
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGG 60

Qy    174 AGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAA 233
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAA 120

Qy    234 GCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAA 293
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAA 180

Qy    294 AGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGC 353
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGC 240

Qy    354 CAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATC 413
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATC 300

Qy    414 TGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGAT 473
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGAT 360

Qy    474 GCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCAC 533
          |||||||||||||||||||||||||||||          ||||||||||||||||||||
Db    361 GCTGCCCGGGGCTCTGCCAGATCCTGAGA----------CGCTTCCCCTCCCTGCCCCAC 410

Qy    534 CCGGGTCCTGTGCTGGCTCCTGCCCCTT----------CAGCCAGGGGTCAGGAGGTGGC 583
          |||||||||||||||||||||||||||||          ||||||||||||||||||||
Db    411 CCGGGTCCTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGC 470

Qy    584 TCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCC 643
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    471 TCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCC 530

Qy    644 CTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGT 703
          |||||||||||||||||||||||||||||| |||| ||||||||||||||||||||||||
Db    531 CTTGGGCTGAGCACCAAGACCTTGAACCTTTCTTGTGTTACCTTTTTTCCAAATAACAGT 590

Qy    704 TGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGTGG 741
          |||||||||||||||| |   | ||||||   | |
Db    591 TGGGAGAAATATCAATGANACATTTGGGGTGACNGNTG 628
```

```
RESULT  33
BI224994
LOCUS       BI224994                659 bp    mRNA    linear   EST 11-JUL-2001
DEFINITION  602949185F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:5093493 5',
            mRNA sequence.
ACCESSION   BI224994
VERSION     BI224994.1  GI:14678438
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1860  row: j  column: 22
            High quality sequence stop: 659.
FEATURES             Location/Qualifiers
     source          1. .659
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5093493"
                     /clone_lib="NIH_MGC_8"
                     /tissue_type="Burkitt lymphoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      156 a    154 c    201 g    148 t
ORIGIN

  Query Match             46.6%;  Score 558.4;  DB 10;  Length 659;
  Best Local Similarity   95.2%;  Pred. No. 2.4e-81;
  Matches  638;  Conservative    0;  Mismatches    6;  Indels   26;  Gaps    5;

Qy    255 GCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACT 314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACT 60

Qy    315 ACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGA 374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGA 120

Qy    375 ACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGC 434
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGC 180

Qy    435 TCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGA 494
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGA 240

Qy    495 TCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCT 554
          ||||||||          ||||||||||||||||||||||||||||||||||||||||||
Db    241 TCCTGAGA----------CGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCT 290

Qy    555 GCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGC 604
          |||||||          ||||||||||||||||||||||||||||||||||||||||||
```

```
Db    291 GCCCCTTCCTGCTTTTGCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGC 350

Qy    605 AGAAGCCCTTT---CCTGTTGGTGTCCCAGCACATGGAGCCCCTT--GGGCTGAGCACCA 659
          ||||| || ||   |  |||||||||||||||||||||||||| |  ||||||||||||
Db    351 AGAAGGCCCTTGTCCTGGTTGGTGTCCCAGCACATGGAGCCCCATTGGGGCTGAGCACCA 410

Qy    660 AGACCTTGAACCTTTTTTGTTTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAAT 719
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    411 AGACCTTGAACCTTTTTTG-TTTACCTTTTTTCCAAATAACAGTTGGGAGAAATATCAAT 469

Qy    720 GAAATTCTGGGGGTGGGGGTGGGGCTGAAAGGGTGGGGTGGGAGATATGGAGGAGTATGA 779
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    470 GAAATTCTGGGGGTGGGGGTGGGGTTGAAAGGGTGGGGTGGGAGATATGGAGGAGTATGA 529

Qy    780 ATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACGTGGCTGATG 839
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    530 ATTAGGGCTTGGAGTTGGTAAAAACATTCCTGACTATCCTTCTTAACCACGTGGCTGATG 589

Qy    840 TGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTAGTTGAGC 899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    590 TGGGGTAGGTATGAGGGGAAGGAAGTGGAGTAGCCTAATGAAAAGGGGTTCTAGTTGAGC 649

Qy    900 TCTGTAGATA 909
          ||||||||||
Db    650 TCTGTAGATA 659
```

```
RESULT  34
BG248926
LOCUS       BG248926                638 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  602361363F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4469943 5',
            mRNA sequence.
ACCESSION   BG248926
VERSION     BG248926.1  GI:12758742
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 638)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10286  row: m  column: 16
            High quality sequence stop: 636.
FEATURES             Location/Qualifiers
     source          1. .638
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4469943"
                     /clone_lib="NIH_MGC_89"
                     /tissue_type="hypernephroma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      154 a    199 c    175 g    110 t
ORIGIN
```

```
  Query Match           46.4%;  Score 556;  DB 10;  Length 638;
  Best Local Similarity  94.8%;  Pred. No. 6.1e-81;
  Matches  607;  Conservative    0;  Mismatches   30;  Indels    3;  Gaps    3;

Qy     31 GTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCCCTTACACTCCA 90
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCCCTTACACTCCA 60

Qy     91 CAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGG 150
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     61 CAGTCCCGGTCCCGCCATGT-CCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGG 119

Qy    151 CGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGC 210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    120 CGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGC 179

Qy    211 AAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCAT 270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 AAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCAT 239

Qy    271 GAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGC 330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    240 GAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGC 299

Qy    331 CAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGA 390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    300 CAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGA 359

Qy    391 TCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCT 450
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    360 TCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCT 419

Qy    451 TGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTT 510
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    420 TGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGAC-GCTTCC 478

Qy    511 GCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTCAGCCAGGG 570
          |  | || ||  || | ||      | |   |||    ||||||   | ||||||||
Db    479 CCTCCCTCGCCCACCGGGTCCTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGG 538

Qy    571 GTCAGGAGGT-GGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCC 629
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    539 GTCAGGAGGTCGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCC 598

Qy    630 AGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAA 669
          ||||||||||||||||||||||||||||||||||||||||
Db    599 AGCACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAA 638
```

```
RESULT  35
BG748595
LOCUS       BG748595                682 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602706289F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842910 5',
            mRNA sequence.
ACCESSION   BG748595
VERSION     BG748595.1  GI:14059248
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 682)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1678  row: a  column: 23
```

```
           High quality sequence stop: 670.
FEATURES             Location/Qualifiers
     source          1. .682
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4842910"
                     /clone_lib="NIH_MGC_43"
                     /tissue_type="normal pigmented retinal epithelium"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library. |"
BASE COUNT      166 a    162 c    209 g    145 t
ORIGIN


  Query Match             45.9%;  Score 550.8;  DB 10;  Length 682;
  Best Local Similarity   95.7%;  Pred. No. 4.1e-80;
  Matches  595;  Conservative    0;  Mismatches    7;  Indels   20;  Gaps    2;

Qy    162 AGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGG 221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60 AGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGG 119

Qy    222 CCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCC 281
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    120 CCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCC 179

Qy    282 AGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGA 341
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 AGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGA 239

Qy    342 ATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCA 401
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    240 ATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCA 299

Qy    402 CCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCC 461
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    300 CCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCC 359

Qy    462 AAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCC 521
          ||||||||||||||||||||||||||||||||||||||||||          |||||||||
Db    360 AAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGA----------CGCTTCCCC 409

Qy    522 TCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCC----------TTCAGCCAGGGG 571
          ||||||||||||||||||||||||||||||||||||||          ||    |  |||
Db    410 TCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCCTTCCTGCTTTTGCAGCCAGGG 469

Qy    572 TCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAG 631
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    470 TCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAG 529

Qy    632 CACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTT 691
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    530 CACATGGAGCCCCTTGGGCTGAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTT 589

Qy    692 CCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGTGGGGCTGAAAGG 751
          ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    590 CCAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGTGGGGTTGAAAGG 649

Qy    752 GTGGGGTGGGAGATATGGAGGA 773
          ||||||||||||||||||||||
Db    650 GTGGGGTGGGAGATATGGAGGA 671


RESULT  36
BI255306
LOCUS       BI255306                614 bp    mRNA    linear   EST 17-JUL-2001
DEFINITION  602977333F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5122089 5',
            mRNA sequence.
ACCESSION   BI255306
VERSION     BI255306.1  GI:14808559
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 614)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11298  row: b  column: 10
            High quality sequence stop: 604.
FEATURES             Location/Qualifiers
     source          1. .614
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5122089"
                     /clone_lib="NIH_MGC_12"
                     /tissue_type="cervical carcinoma cell line"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.4 kb. Library prepared by Life
                     Technologies."
BASE COUNT      149 a    191 c    170 g    104 t
ORIGIN


  Query Match             45.9%;  Score 550.6;  DB 10;  Length 614;
  Best Local Similarity   95.5%;  Pred. No. 4.6e-80;
  Matches  595;  Conservative    0;  Mismatches    9;  Indels   19;  Gaps    2;

Qy     31 GTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCCCTTACACTCCA 90
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCCCTTACACTCCA 60

Qy     91 CAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGG 150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGG 120

Qy    151 CGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGC 210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGC 180

Qy    211 AAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCAT 270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCAT 240

Qy    271 GAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGC 330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGC 300

Qy    331 CAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGA 390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGA 360

Qy    391 TCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCT 450
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCACATCCCCAGCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCT 420

Qy    451 TGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTT 510
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||          | |
Db    421 TGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGA---------ATCCTGA 471

Qy    511 GCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTT--------- 561
          | ||||||||||||||||||| ||||||||||||||||||||||||||
Db    472 GACGCTTCCCCTCCCTGCCCCAGCCGGGTCCTGTGCTGGCTCCTGCCCCTTCCTGCTTTT 531

Qy    562 -CAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGT 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    532 GCAGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGT 591

Qy    621 TGGTGTCCCAGCACATGGAGCCC 643
          |||||||||||||||||||||||
Db    592 AGGTGTCCCAGCACATGGAGCCC 614


RESULT  37
AW965834
LOCUS       AW965834                636 bp    mRNA    linear   EST 01-JUN-2000
DEFINITION  EST377907 MAGE resequences, MAGI Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW965834
VERSION     AW965834.1  GI:8155670
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 636)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 219
            Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source          1. .636
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGI"
                     /note="Vector: pBluescriptSKm"
BASE COUNT      151 a    202 c    170 g    109 t      4 others
ORIGIN


  Query Match             45.9%;  Score 549.8;  DB 9;  Length 636;
  Best Local Similarity   93.5%;  Pred. No. 6.2e-80;
  Matches  601;  Conservative    0;  Mismatches   21;  Indels   21;  Gaps    2;

Qy     18 AGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGC 77
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGC 60

Qy     78 CCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTG 137
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTG 120

Qy    138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 180

Qy    198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 240

Qy    258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 300

Qy    318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 360

Qy    378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 420

Qy    438 TCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCC 497
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCC 480

Qy    498 TGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCC 557
          |||||          ||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGAGA----------CGCTTCCCCTCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCC 530

Qy    558 CCTTC-----------AGCCAGGGGTCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAG 606
          |||||           |   ||||||||||||||   |||||||| ||||||||
Db    531 CCTTCCTGCTTTTGCAGCCAGNGGGTCAGGAGGTGGCCTCGGTGTGGGCCTGAGAGGCAN 590

Qy    607 AAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGG 649
          ||  || ||||   | ||||||| ||||||||||||| | ||
Db    591 AANNCCCTTCCCTGTTGTGTCCCAACACATGGAGCCCCCTTGG 633


RESULT  38
BF982805
LOCUS       BF982805                983 bp    mRNA    linear   EST 23-JAN-2001
DEFINITION  602304840F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4396280 5',
            mRNA sequence.
ACCESSION   BF982805
VERSION     BF982805.1  GI:12385617
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 983)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10094  row: p  column: 09
            High quality sequence stop: 646.
FEATURES             Location/Qualifiers
     source          1. .983
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4396280"
                     /clone_lib="NIH_MGC_88"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      297 a    239 c    288 g    159 t
ORIGIN


  Query Match             45.6%;  Score 546.4;  DB 10;  Length 983;
```

```
  Best Local Similarity   91.0%;  Pred. No. 1.9e-79;
  Matches  655;  Conservative    0;  Mismatches   41;  Indels   24;  Gaps    6;

Qy     163 GGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGC 222
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8 GAACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGC 67

Qy     223 CAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCA 282
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68 CAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCA 127

Qy     283 GAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAA 342
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     128 GAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAA 187

Qy     343 TAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCAC 402
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     188 TAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCAC 247

Qy     403 CCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCA 462
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248 CCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCA 307

Qy     463 AGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCT 522
           ||||||||||||||||||||||||||||||||||||||||||||          ||||||||||
Db     308 AGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGA----------CGCTTCCCCT 357

Qy     523 CCCTGCCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCCTTC---------AGCCAGGGGTC 573
           ||||| |||||||||||||||||||||||||||||||||||          | |||||
Db     358 CCCTG-CCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGTC 416

Qy     574 AGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCA 633
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     417 AGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCA 476

Qy     634 CATGGAGCCCCTTGGGCTGAGC-ACCAAGACCTTGAACCTTTTTTGTTTTACCTTTTTTC 692
           |||||||||||||||||||| |||||||||||||||||||| ||||||||||||||||
Db     477 CATGGAGCCCCTTGGGCTGAGCAACCAAGACCTTGAACCTTTTTAGTTTTACCTTTTTTC 536

Qy     693 CAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTG-GGGGTGGGGCTGAAAGG 751
           |||||||||||||||||||||||||||||||||||||||| | ||||||| ||||| |
Db     537 CAAATAACAGTTGGGAGAAATATCAATGAAATTCTGGGGGTGAGAGGTGGGGTTGAAAAG 596

Qy     752 GTGGGGTGGG--AGATATGGAGGAGTATGAATTAGGGCTTGGAGTTGGTAAAAACATTCC 809
           ||| ||||||       ||| |||||||||||||||||||||| |  |||| |||||
Db     597 GTGAGGTGGGGAGATATTGGCGGAGTATGAATTAGGGCTTGGAGTGGTAAAAACCATTCA 656

Qy     810 TGACTATCCTTCTTAACCACGTGGCTGATGTGGGGTAGGTATGAGGGGAAGGAAGTGGAG 869
           |||||||||||||||| ||| | ||| ||||| | |       |  || |||||   ||
Db     657 TGACTATCCTTCTTAACAACGGGCCTGCTGTGGAGGAAAGGACTGACGAGGGAAGGAAAG 716


RESULT  39
BG576751
LOCUS       BG576751                 707 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  602599070F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4707976 5',
            mRNA sequence.
ACCESSION   BG576751
VERSION     BG576751.1  GI:13584404
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 707)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10578  row: k  column: 17
            High quality sequence stop: 613.
FEATURES             Location/Qualifiers
     source          1. .707
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4707976"
                     /clone_lib="NIH_MGC_87"
                     /tissue_type="mammary adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.383 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      172 a    213 c    194 g    128 t
ORIGIN


  Query Match           45.1%;  Score 540.6;  DB 10;  Length 707;
  Best Local Similarity   92.3%;  Pred. No. 1.9e-78;
  Matches  643;  Conservative    0;  Mismatches   29;  Indels   25;  Gaps    6;

Qy       3 GACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCC-GTTTCCCCTCCCCCTTCCCTTCTCC 61
           |||  | |    |||||||||||||||||||||| ||| ||||||||||||||||||||
Db       3 GACGCGTGGGTCAGCAGCAGCGCAGGTTGTCCCCGGTTGCCCCTCCCCCTTCCCTTCTCC 62

Qy      62 GGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAA 121
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63 GGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAA 122

Qy     122 GAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAA 181
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123 GAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAA 182

Qy     182 GGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGA 241
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     183 GGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGA 242

Qy     242 CAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTT 301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     243 CAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTT 302

Qy     302 GACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCA 361
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     303 GACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCA 362

Qy     362 GGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAG 421
           |||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db     363 GGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCGACCCACAGGATCTGCCCCAG 422

Qy     422 AGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCG 481
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     423 AGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCCAAGTTGAATGATGCTGCCCG 482

Qy     482 GGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCCTCCCTGCCCCACCCGGGTCC 541
           |||||||||||||||||||||||||||||||||| |||||||||  ||||||||||||||
Db     483 GGGCTCTGCCAGATCCTGAGAC----------GCTTCCCTCCCTGACCCACCCGGGTCC 531

Qy     542 TGTGCTGGCTCCTGCCCCTT----------CAGCCAGGGGTCAGGAGGTGGCTCGGGTGT 591
           ||||||||||||||||||||          |||||||||||| || ||||||||||||||
Db     532 TGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGGTCCGGCGGTGGCTCGGGTGT 591

Qy     592 GGGCTGGAGAGGCAGAAGCCCTTTCCTGTTGGTGTCCCAGCACATGGAGCCCCTTGGGCT 651
           | |||||||||||||||||||||||||||| ||||| ||||||||| ||| |||| |||||
Db     592 GAGCTGGAGAGGCAGAAGCCCTTTCCTGTAGGTGT-CCAGCACATTGAG-CCCTAGGGCT 649

Qy     652 GAGCACCAAGACCTTGAACCTTTTTTGTTTTACCTTT 688
           ||||| ||||||||||||||||  |  ||     |||
```

```
Db    650 GAGCA-CAAGACCTTGAACCTTACTGTTTACCTTTTT 685


RESULT  40
BM018285
LOCUS       BM018285                718 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  603645940F1 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5427705 5',
            mRNA sequence.
ACCESSION   BM018285
VERSION     BM018285.1  GI:16532639
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 718)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1893  row: p  column: 10
            High quality sequence stop: 718.
FEATURES             Location/Qualifiers
     source          1. .718
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5427705"
                     /clone_lib="NIH_MGC_98"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      179 a    164 c    228 g    147 t
ORIGIN


  Query Match             44.6%;  Score 535.2;  DB 10;  Length 718;
  Best Local Similarity   93.0%;  Pred. No. 1.4e-77;
  Matches  668;  Conservative    0;  Mismatches   18;  Indels   32;  Gaps    9;

Qy    162 AGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGG 221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13 AGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGG 72

Qy    222 CCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCC 281
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     73 CCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCC 132

Qy    282 AGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGA 341
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    133 AGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGA 192

Qy    342 ATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCA 401
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    193 ATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCA 252

Qy    402 CCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCC 461
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    253 CCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGTGGCC 312

Qy    462 AAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGACTGCTTTTGCCGCTTCCCC 521
          ||||||||||||||||||||||||||||||||||||||||||          |||||||||
Db    313 AAGTTGAATGATGCTGCCCGGGGCTCTGCCAGATCCTGAGA----------CGCTTCCCC 362

Qy    522 TCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTT----------CAGCCAGGGG 571
          ||||||||||||||||||||||||||||||||||||||||          ||||||||||
Db    363 TCCCTGCCCCACCCGGGTCCTGTGCTGGCTCCTGCCCCTTCCTGCTTTTGCAGCCAGGGG 422

Qy    572 TCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAA-GCCCTTTCCTGTTGGTGTCCCA 630
          |||||||||||||||||||||||||||||||||||||| |||||| ||||||||||||||
Db    423 TCAGGAGGTGGCTCGGGTGTGGGCTGGAGAGGCAGAAGGCCCTTGCCTGTTGGTGTCCCA 482

Qy    631 GCACATGGAGCCCCTT-GGGCTGAGCACCAAGACCTTGAACCTTTTT-----TGTTTTAC 684
          |||||||||||||||| |||||||||||||||||||     |||     | | |  |
Db    483 GCACATGGAGCCCCTTGGGGCTGAGCACCAAGACCTTGAGAACCTTTGTTGGTTTGTACC 542

Qy    685 CTTTTTTCCAAATAACAGT-TGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGTGGGG 743
           |||||  |||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    543 TTTTTTGGCAAATAACAGTGTGGGAGAAATATCAATGAAATTCTGGGGGTGGGGGTGGGG 602

Qy    744 CTGAAAGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAGGGC-TTGGAGTTGGTAAAA 802
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    603 TGGAAAGGGTGGGGTGGGAGATATGGAGGAGTATGAATTAGGGCGTTGGAGTTGGTAAAA 662

Qy    803 AC-ATTCCTGACTATCCTTCTTAACCACGTGGCTGATGTGGGGTAGGTATGAGGGGAA 859
          || || |||||||||||| |||||   ||||||||||||||||||||||||||||||
Db    663 ACGATGCCTGACTATCCTTGTTAAC--AGTGGCTGATGTGGGGTAGGTATGAGGGGAA 718


Search completed: September 12, 2002, 16:05:46
Job time: 22145 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2002, 09:42:16 ; Search time 4285.42 Seconds
(without alignments)
5386.163 Million cell updates/sec

Title: US-09-824-178-1
Perfect score: 1103
Sequence: 1 GTGACAGGAGCCGAAGCAGC.....AAAAAAAAAAGGGCGGCGCG 1103

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
		%				
1	1093	99.1	1093	9	AF157510	AF157510 Homo sapi
c 2	1033.2	93.7	148578	9	CNS01RIF	AL163171 Human chr
c 3	1033.2	93.7	159846	2	AC069588	AC069588 Homo sapi
c 4	790.8	71.7	176550	9	AL356356	AL356356 Human DNA
c 5	790.8	71.7	207815	2	AC053497	AC053497 Homo sapi
6	738.2	66.9	194304	9	AC016708	AC016708 Homo sapi
7	456	41.3	749	9	HSALPEND	X99906 Homo sapien
8	456	41.3	1199	9	AF157509	AF157509 Homo sapi
9	456	41.3	1208	9	AK001981	AK001981 Homo sapi
10	454.4	41.2	1788	9	HSAJ10966	AJ010966 Homo sapi
11	450.8	40.9	1211	9	BC000436	BC000436 Homo sapi
12	450.6	40.9	69732	2	AC091596	AC091596 Homo sapi
13	450	40.8	1202	9	BC004461	BC004461 Homo sapi
14	431.4	39.1	669	6	AR144151	AR144151 Sequence
15	431.4	39.1	669	6	AR176338	AR176338 Sequence
16	342.2	31.0	98113	2	AC023294	AC023294 Homo sapi
17	334	30.3	366	4	SSA005987	AJ005987 Sus scrof
18	332.4	30.1	366	4	BTA005986	AJ005986 Bos tauru
19	332.4	30.1	366	10	MMA005985	AJ005985 Mus muscu
20	329.2	29.8	366	10	RNA005984	AJ005984 Rattus no
c 21	321	29.1	142023	2	AC092479	AC092479 Mus muscu
c 22	278.6	25.3	89872	9	HSJ822J19	AL050324 Human DNA
c 23	277	25.1	89263	9	AC007124	AC007124 Human gen
24	276.6	25.1	176616	2	AC007449	AC007449 Homo sapi
25	228.2	20.7	105126	2	AC103421	AC103421 Rattus no
26	213.4	19.3	231	4	BTAENSULF	X95958 B.taurus mR
27	187.8	17.0	194799	2	AC096386	AC096386 Rattus no
28	176.2	16.0	339	10	MMA005983	AJ005983 Mus muscu
29	171	15.5	339	10	RNA005982	AJ005982 Rattus no
30	169.4	15.4	339	9	HOS223091	AJ223091 Homo sapi
31	169.4	15.4	1188	4	BOVARPP19	M33618 Bovine cAMP
32	169.4	15.4	1228	9	BC003418	BC003418 Homo sapi
33	169.4	15.4	5171	9	AF084555	AF084555 Homo sapi
34	167.8	15.2	339	4	SSA005981	AJ005981 Sus scrof
35	166.8	15.1	71941	2	AC091874	AC091874 Homo sapi
36	165.4	15.0	2288	9	AK057673	AK057673 Homo sapi
37	159.4	14.5	1961	4	BOVARPP16	M33617 Bovine cAMP
38	156.6	14.2	173169	9	AC068538	AC068538 Homo sapi
39	113.6	10.3	221899	2	AC079548	AC079548 Mus muscu
c 40	100	9.1	70244	2	AC090836	AC090836 Homo sapi
c 41	100	9.1	149081	2	AC025917	AC025917 Homo sapi
c 42	100	9.1	184128	2	AC073438	AC073438 Homo sapi
c 43	100	9.1	219389	2	AC009754	AC009754 Homo sapi
44	74	6.7	246262	2	AL669825	AL669825 Mus muscu
c 45	72.2	6.5	170548	10	AC074041	AC074041 Mus muscu

ALIGNMENTS

RESULT 1
AF157510
LOCUS AF157510 1093 bp mRNA linear PRI 01-JUL-2000
DEFINITION Homo sapiens beta endosulfine mRNA, complete cds.
ACCESSION AF157510
VERSION AF157510.1 GI:8886010
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1093)
AUTHORS Scott,V.E.S., Roch,J.-M., Davis-Taber,R.A., Molinari,E.J., Whiteaker,K.L., Gopalakrishnan,M., Idler,K. and Sullivan,J.P.
TITLE Cloning and molecular characterization of two isoforms of human endosulfine
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1093)
AUTHORS Scott,V.E.S., Roch,J.-M., Whiteaker,K.L. and Sullivan,J.P.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-1999) D47C, Abbott Laboratories, AP10 L08, 100 Abbott Park, Abbott Park, IL 60064, USA
FEATURES Location/Qualifiers

```
     source          1. .1093
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             107. .460
                     /codon_start=1
                     /product="beta endosulfine"
                     /protein_id="AAF80341.1"
                     /db_xref="GI:8886011"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAG"
BASE COUNT      313 a    261 c    260 g    259 t
ORIGIN


  Query Match          99.1%;  Score 1093;  DB 9;  Length 1093;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1093;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCC 480

Qy     481 CCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCA 540

Qy     541 GATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGT 600

Qy     601 AATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACA 660

Qy     661 CTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAA 720

Qy     721 ATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 ATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCC 780

Qy     781 CAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAAC 840

Qy     841 CTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 CTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCT 900

Qy     901 GTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAA 960

Qy     961 AGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 AGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTAT 1020

Qy    1021 CTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACATTATTCACAAAA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACATTATTCACAAAA 1080

Qy    1081 AAAAAAAAAAAAA 1093
          |||||||||||||
Db    1081 AAAAAAAAAAAAA 1093


RESULT   2
CNS01RIF/c
LOCUS       CNS01RIF              148578 bp    DNA     linear   PRI 26-APR-2001
DEFINITION  Human chromosome 14 DNA sequence BAC R-406A9 of library RPCI-11
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL163171
VERSION     AL163171.4  GI:11611169
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 148578)
  AUTHORS   Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
  TITLE     Sequencing of the human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148578)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Dec 9, 2000 this sequence version replaced gi:9453784.
            -------------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            --------------
            The following BAC sequence is oriented from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : C-2007P7
            Downstream BAC (overlapping the SP6 end) : R-11K13 (AC=AL355095)
            -------------- Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 7.83x in Q20 bases; sum-of-contigs
            --------------
            ------------------------
            Overall quality chart :
            Range     :   bases
            0         :
            1  -  9   :       5
            10 - 19   :      94
            20 - 29   :     158
            30 - 39   :     446
            40 - 49   :    1989
            50 - 59   :    3410
            60 - 69   :    6039
```

```
           70 - 79  :   17886
           80 - 89  :   47912
           90 - 99  :   70639
           ----------------------
           Percentage of bases with a quality value >= 40 : 99 %.
FEATURES             Location/Qualifiers
     source          1. .148578
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="R-406A9"
                     /clone_lib="RPCI-11"
     STS             56891. .57046
                     /note="matching EMBL:R48962
                     RHdb:RH44475
                     dbSTS:STS37544
                     Identified using the e-PCR software (G. Schuler)"
     STS             139692. .139819
                     /note="matching EMBL:G04746
                     RHdb:RH53610
                     RHdb:RH3734
                     dbSTS:STS25994
                     Identified using the e-PCR software (G. Schuler)"
BASE COUNT    46908 a  26225 c  26858 g  48587 t
ORIGIN


  Query Match           93.7%;  Score 1033.2;  DB 9;  Length 148578;
  Best Local Similarity   97.8%;  Pred. No. 1.7e-292;
  Matches 1058;  Conservative    0;  Mismatches   23;  Indels    1;  Gaps    1;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           |||| ||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   59279 GTGATAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCTCTCCCCCTTCCCTTCTC 59220

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   59219 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAGCA 59160

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           |||||||||||||||| |||||| ||||||||||||||||||||||||||||||||||||
Db   59159 AGAAGAAGAGAACCCTGCAGAGGAGGCCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 59100

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   59099 AGGTATTCTCCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 59040

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   59039 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 58980

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   58979 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 58920

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
           |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   58919 AGGACCAGACAAGAACCTGGTGACTGGTGATCACAT-CCCACCCCACAGGATCTGCCCCA 58861

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCC 480
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   58860 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGACCCCCTCTCCTCC 58801

Qy     481 CCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCA 540
           |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db   58800 CCTTCCTCAACCATTGGACTTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCA 58741

Qy     541 GATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGT 600
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   58740 GATCTTCTCAGCTTGCTAGCCAAAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGT 58681

Qy     601 AATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACA 660
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   58680 AATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTCTTTCGTTAGATTGACACTTCCACACA 58621

Qy     661 CTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAA 720
           |||||||||||||||||||||||| |||||||||| |||||||||||||||||| ||||||||||
Db   58620 CTCCCTGTAGTCCAGGTAGGGCCCAGAAATAGGAAAGGCTAGGATTGGACAATGCTGCAA 58561

Qy     721 ATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCC 780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   58560 ATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCC 58501

Qy     781 CAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAAC 840
           |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   58500 CAAAATTGATTAGGCTGAGCCTTGGGAATAATTTGGCAGGTTTAACATCCCAAGGCTAAC 58441

Qy     841 CTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCT 900
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   58440 CTAACATAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCT 58381

Qy     901 GTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAA 960
           ||||||| |||||||||||||||||||||||||||  ||||||||||||||||||||||
Db   58380 GTCCCTTCGGCTGCTATTGCTTCATGTTTCCATTATGGCGAGTTTAGAGAATCCTTAAAA 58321

Qy     961 AGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTAT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   58320 AGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTAT 58261

Qy    1021 CTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACATTATTCACAAAA 1080
           |||||||||||||||||| || |||| |||| |||||||||||||||||||||||| ||
Db   58260 CTCTTACAGTTTGCTCTGCCTTTCAAGTAATGTAAAGATTGATGAACATTATTCACAGAA 58201

Qy    1081 AA 1082
           ||
Db   58200 AA 58199


RESULT   3
AC069588/c
LOCUS       AC069588              159846 bp    DNA     linear   HTG 04-OCT-2000
DEFINITION  Homo sapiens chromosome 14 clone RP11-340E1 map 14, WORKING DRAFT
            SEQUENCE, 22 unordered pieces.
ACCESSION   AC069588
VERSION     AC069588.3  GI:10567972
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159846)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 14, clone RP11-340E1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 159846)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
```

```
TITLE     Direct Submission
JOURNAL   Submitted (04-JUN-2000) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Oct 4, 2000 this sequence version replaced gi:8569195.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          -------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
              Center project name: L8786
              Center clone name: 340_E_1
          -------------- Summary Statistics
              Sequencing vector: M13; M77815; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 150080 bases at least Q40
              Consensus quality: 154891 bases at least Q30
              Consensus quality: 156760 bases at least Q20
              Insert size: 157000; agarose-fp
              Insert size: 157746; sum-of-contigs
              Quality coverage: 4.3 in Q20 bases; agarose-fp
              Quality coverage: 4.3 in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 22 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    11674: contig of 11674 bp in length
          *    11675 11774: gap of      100 bp
          *    11775    12783: contig of 1009 bp in length
          *    12784 12883: gap of      100 bp
          *    12884    14380: contig of 1497 bp in length
          *    14381 14480: gap of      100 bp
          *    14481    15987: contig of 1507 bp in length
          *    15988 16087: gap of      100 bp
          *    16088    20096: contig of 4009 bp in length
          *    20097 20196: gap of      100 bp
          *    20197    24318: contig of 4122 bp in length
          *    24319 24418: gap of      100 bp
          *    24419    28596: contig of 4178 bp in length
          *    28597 28696: gap of      100 bp
          *    28697    32856: contig of 4160 bp in length
          *    32857 32956: gap of      100 bp
          *    32957    37590: contig of 4634 bp in length
          *    37591 37690: gap of      100 bp
          *    37691    43359: contig of 5669 bp in length
          *    43360 43459: gap of      100 bp
          *    43460    49215: contig of 5756 bp in length
          *    49216 49315: gap of      100 bp
          *    49316    55482: contig of 6167 bp in length
          *    55483 55582: gap of      100 bp
          *    55583    61888: contig of 6306 bp in length
          *    61889 61988: gap of      100 bp
          *    61989    70075: contig of 8087 bp in length
          *    70076 70175: gap of      100 bp
          *    70176    79049: contig of 8874 bp in length
          *    79050 79149: gap of      100 bp
          *    79150    87232: contig of 8083 bp in length
          *    87233 87332: gap of      100 bp
          *    87333    98465: contig of 11133 bp in length
          *    98466 98565: gap of      100 bp
          *    98566   116653: contig of 18088 bp in length
          *   116654 116753: gap of      100 bp
          *   116754   127734: contig of 10981 bp in length
          *   127735 127834: gap of      100 bp
          *   127835   139387: contig of 11553 bp in length
          *   139388 139487: gap of      100 bp
          *   139488   150540: contig of 11053 bp in length
          *   150541 150640: gap of      100 bp
          *   150641   159846: contig of 9206 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159846
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14"
                     /clone="RP11-340E1"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .11674
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    11775. .12783
                     /note="assembly_fragment"
     misc_feature    12884. .14380
                     /note="assembly_fragment"
     misc_feature    14481. .15987
                     /note="assembly_fragment"
     misc_feature    16088. .20096
                     /note="assembly_fragment"
     misc_feature    20197. .24318
                     /note="assembly_fragment"
     misc_feature    24419. .28596
                     /note="assembly_fragment"
     misc_feature    28697. .32856
                     /note="assembly_fragment"
     misc_feature    32957. .37590
                     /note="assembly_fragment"
     misc_feature    37691. .43359
                     /note="assembly_fragment"
     misc_feature    43460. .49215
                     /note="assembly_fragment"
     misc_feature    49316. .55482
                     /note="assembly_fragment"
     misc_feature    55583. .61888
                     /note="assembly_fragment"
     misc_feature    61989. .70075
                     /note="assembly_fragment"
     misc_feature    70176. .79049
                     /note="assembly_fragment"
     misc_feature    79150. .87232
                     /note="assembly_fragment"
     misc_feature    87333. .98465
                     /note="assembly_fragment"
     misc_feature    98566. .116653
                     /note="assembly_fragment"
     misc_feature    116754. .127734
                     /note="assembly_fragment"
     misc_feature    127835. .139387
                     /note="assembly_fragment"
     misc_feature    139488. .150540
                     /note="assembly_fragment"
     misc_feature    150641. .159846
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    49405 a  27703 c  27568 g  53067 t   2103 others
ORIGIN


  Query Match             93.7%;  Score 1033.2;  DB 2;  Length 159846;
  Best Local Similarity   97.8%;  Pred. No. 1.7e-292;
  Matches 1058;  Conservative    0;  Mismatches   23;  Indels    1;  Gaps    1;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           |||| |||||||||||||||||||||||||||||||||| |||||||||||||||||
Db  119501 GTGATAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCTCTCCCCCTTCCCTTCTC 119442
```

```
Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db 119441 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAGCA 119382

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           |||||||||||||||| |||||| ||||||||||||||||||||||||||||||||||||
Db 119381 AGAAGAAGAGAACCCTGCAGAGGAGGCCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 119322

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 119321 AGGTATTCTCCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 119262

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 119261 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 119202

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 119201 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 119142

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db 119141 AGGACCAGACAAGAACCTGGTGACTGGTGATCACAT-CCCACCCCACAGGATCTGCCCCA 119083

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCC 480
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db 119082 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGACCCCCTCTCCTCC 119023

Qy     481 CCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCA 540
           |||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db 119022 CCTTCCTCAACCATTGGACTTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCA 118963

Qy     541 GATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGT 600
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db 118962 GATCTTCTCAGCTTGCTAGCCAAAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGT 118903

Qy     601 AATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACA 660
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db 118902 AATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTCTTTCGTTAGATTGACACTTCCACACA 118843

Qy     661 CTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAA 720
           |||||||||||||||||||||| |||||||||||||||||||||||| ||||||||||||
Db 118842 CTCCCTGTAGTCCAGGTAGGGCCCAGAAATAGGAAAGGCTAGGATTGGACAATGCTGCAA 118783

Qy     721 ATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 118782 ATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCC 118723

Qy     781 CAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAAC 840
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db 118722 CAAAATTGATTAGGCTGAGCCTTGGGAATAATTTGGCAGGTTTAACATCCCAAGGCTAAC 118663

Qy     841 CTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCT 900
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 118662 CTAACATAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCT 118603

Qy     901 GTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAA 960
           ||||||| |||||||||||||||||||||||||||||  ||||||||||||||||||||
Db 118602 GTCCCTTCGGCTGCTATTGCTTCATGTTTCCATTATGGCGAGTTTAGAGAATCCTTAAAA 118543

Qy     961 AGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTAT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 118542 AGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTAT 118483

Qy    1021 CTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACATTATTCACAAAA 1080
           |||||||||||||||| || |||| |||| |||||||||||||||||||||||||| ||
Db 118482 CTCTTACAGTTTGCTCTGCCTTTCAAGTAATGTAAAGATTGATGAACATTATTCACAGAA 118423

Qy    1081 AA 1082
           ||
Db 118422 AA 118421
```

RESULT 4
AL356356/c
LOCUS AL356356 176550 bp DNA linear PRI 30-JAN-2002
DEFINITION Human DNA sequence from clone RP11-54A4 on chromosome 1, complete sequence.
ACCESSION AL356356
VERSION AL356356.17 GI:18476570
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 176550)
AUTHORS Heath,P.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Feb 1, 2002 this sequence version replaced gi:18250748.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr1
RP11-54A4 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-54A4 It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone RP11-54A4 is at 1 in this sequence. The true left end of clone RP11-363I22 is at 174551 in this sequence.

```
FEATURES             Location/Qualifiers
     source          1. .176550
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-54A4"
                     /clone_lib="RPCI-11.1"
     misc_feature    50307. .50323
                     /note="Single clone region. Reads generated from a
                     transposon library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    123022. .123038
                     /note="Sequence from uni-directional primer reads and dGTP
                     big dye terminator reads only."
BASE COUNT    45999 a  42218 c  42262 g  46071 t
ORIGIN
```

Query Match 71.7%; Score 790.8; DB 9; Length 176550;
Best Local Similarity 99.7%; Pred. No. 4.6e-221;
Matches 792; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy     289 GCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCA 348
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153071 GCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCA 153012

Qy    349 GCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACA 408
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 153011 GCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACA 152952

Qy    409 GGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCC 468
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152951 GGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCC 152892

Qy    469 CCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATG 528
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152891 CCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATG 152832

Qy    529 GGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGC 588
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152831 GGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGC 152772

Qy    589 TGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGA 648
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152771 TGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGA 152712

Qy    649 CACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGG 708
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db 152711 CACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCCAGAAATAGGAAAGGCTAGGATTGG 152652

Qy    709 ATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGG 768
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152651 ATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGG 152592

Qy    769 GAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACAT 828
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152591 GAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACAT 152532

Qy    829 CCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCT 888
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152531 CCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCT 152472

Qy    889 TCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGA 948
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152471 TCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGA 152412

Qy    949 GAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCC 1008
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152411 GAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCC 152352

Qy   1009 ATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACA 1068
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152351 ATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACA 152292

Qy   1069 TTATTCACAAAAAA 1082
          ||||||||| ||||
Db 152291 TTATTCACAGAAAA 152278


RESULT   5
AC053497/c
LOCUS       AC053497               207815 bp    DNA     linear   HTG 13-AUG-2000
DEFINITION  Homo sapiens chromosome 1 clone RP11-243G22, WORKING DRAFT
            SEQUENCE, 37 unordered pieces.
ACCESSION   AC053497
VERSION     AC053497.4  GI:9797901
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 207815)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 207815)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8439979.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0243G22
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 192323 bases at least Q40
            Consensus quality: 196320 bases at least Q30
            Consensus quality: 198661 bases at least Q20
            Insert size: 204000; agarose-fp
            Insert size: 204215; sum-of-contigs
            Quality coverage: 6.36 in Q20 bases; agarose-fp
            Quality coverage: 6.31 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 37 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1381: contig of 1381 bp in length
            *     1382     1481: gap of unknown length
            *     1482     2602: contig of 1121 bp in length
            *     2603     2702: gap of unknown length
            *     2703     4013: contig of 1311 bp in length
            *     4014     4113: gap of unknown length
            *     4114     5331: contig of 1218 bp in length
            *     5332     5431: gap of unknown length
            *     5432     6622: contig of 1191 bp in length
            *     6623     6722: gap of unknown length
            *     6723     8166: contig of 1444 bp in length
            *     8167     8266: gap of unknown length
            *     8267     9674: contig of 1408 bp in length
            *     9675     9774: gap of unknown length
            *     9775    10943: contig of 1169 bp in length
            *    10944    11043: gap of unknown length
            *    11044    12198: contig of 1155 bp in length
            *    12199    12298: gap of unknown length
            *    12299    14193: contig of 1895 bp in length
            *    14194    14293: gap of unknown length
            *    14294    15717: contig of 1424 bp in length
            *    15718    15817: gap of unknown length
            *    15818    16935: contig of 1118 bp in length
            *    16936    17035: gap of unknown length
            *    17036    18375: contig of 1340 bp in length
            *    18376    18475: gap of unknown length
            *    18476    19962: contig of 1487 bp in length
            *    19963    20062: gap of unknown length
            *    20063    21169: contig of 1107 bp in length
            *    21170    21269: gap of unknown length
            *    21270    22452: contig of 1183 bp in length
            *    22453    22552: gap of unknown length
            *    22553    24582: contig of 2030 bp in length
            *    24583    24682: gap of unknown length
            *    24683    26383: contig of 1701 bp in length
            *    26384    26483: gap of unknown length
            *    26484    27710: contig of 1227 bp in length
            *    27711    27810: gap of unknown length
```

```
            *    27811    29195: contig of 1385 bp in length
            *    29196    29295: gap of unknown length
            *    29296    30800: contig of 1505 bp in length
            *    30801    30900: gap of unknown length
            *    30901    32555: contig of 1655 bp in length
            *    32556    32655: gap of unknown length
            *    32656    34653: contig of 1998 bp in length
            *    34654    34753: gap of unknown length
            *    34754    37122: contig of 2369 bp in length
            *    37123    37222: gap of unknown length
            *    37223    39506: contig of 2284 bp in length
            *    39507    39606: gap of unknown length
            *    39607    41220: contig of 1614 bp in length
            *    41221    41320: gap of unknown length
            *    41321    42978: contig of 1658 bp in length
            *    42979    43078: gap of unknown length
            *    43079    45773: contig of 2695 bp in length
            *    45774    45873: gap of unknown length
            *    45874    55056: contig of 9183 bp in length
            *    55057    55156: gap of unknown length
            *    55157    67059: contig of 11903 bp in length
            *    67060    67159: gap of unknown length
            *    67160    80260: contig of 13101 bp in length
            *    80261    80360: gap of unknown length
            *    80361    96528: contig of 16168 bp in length
            *    96529    96628: gap of unknown length
            *    96629   114529: contig of 17901 bp in length
            *   114530   114629: gap of unknown length
            *   114630   133673: contig of 19044 bp in length
            *   133674   133773: gap of unknown length
            *   133774   153791: contig of 20018 bp in length
            *   153792   153891: gap of unknown length
            *   153892   176079: contig of 22188 bp in length
            *   176080   176179: gap of unknown length
            *   176180   207815: contig of 31636 bp in length.
FEATURES             Location/Qualifiers
     source          1. .207815
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-243G22"
     misc_feature    1. .1381
                     /note="assembly_name:Contig13"
     misc_feature    1482. .2602
                     /note="assembly_name:Contig15"
     misc_feature    2703. .4013
                     /note="assembly_name:Contig21"
     misc_feature    4114. .5331
                     /note="assembly_name:Contig25"
     misc_feature    5432. .6622
                     /note="assembly_name:Contig33"
     misc_feature    6723. .8166
                     /note="assembly_name:Contig37"
     misc_feature    8267. .9674
                     /note="assembly_name:Contig38"
     misc_feature    9775. .10943
                     /note="assembly_name:Contig40"
     misc_feature    11044. .12198
                     /note="assembly_name:Contig42"
     misc_feature    12299. .14193
                     /note="assembly_name:Contig45"
     misc_feature    14294. .15717
                     /note="assembly_name:Contig47"
     misc_feature    15818. .16935
                     /note="assembly_name:Contig50"
     misc_feature    17036. .18375
                     /note="assembly_name:Contig51"
     misc_feature    18476. .19962
                     /note="assembly_name:Contig52"
     misc_feature    20063. .21169
                     /note="assembly_name:Contig53"
     misc_feature    21270. .22452
                     /note="assembly_name:Contig54"
     misc_feature    22553. .24582
                     /note="assembly_name:Contig56"
     misc_feature    24683. .26383
                     /note="assembly_name:Contig57"
     misc_feature    26484. .27710
                     /note="assembly_name:Contig58"
     misc_feature    27811. .29195
                     /note="assembly_name:Contig59"
     misc_feature    29296. .30800
                     /note="assembly_name:Contig60"
     misc_feature    30901. .32555
                     /note="assembly_name:Contig63"
     misc_feature    32656. .34653
                     /note="assembly_name:Contig64"
     misc_feature    34754. .37122
                     /note="assembly_name:Contig65"
     misc_feature    37223. .39506
                     /note="assembly_name:Contig66"
     misc_feature    39607. .41220
                     /note="assembly_name:Contig67"
     misc_feature    41321. .42978
                     /note="assembly_name:Contig68"
     misc_feature    43079. .45773
                     /note="assembly_name:Contig69
                     clone_end:T7
                     vector_side:left"
     misc_feature    45874. .55056
                     /note="assembly_name:Contig70"
     misc_feature    55157. .67059
                     /note="assembly_name:Contig71"
     misc_feature    67160. .80260
                     /note="assembly_name:Contig72"
     misc_feature    80361. .96528
                     /note="assembly_name:Contig73"
     misc_feature    96629. .114529
                     /note="assembly_name:Contig74"
     misc_feature    114630. .133673
                     /note="assembly_name:Contig75"
     misc_feature    133774. .153791

  Query Match             71.7%;  Score 790.8;  DB 2;  Length 207815;
  Best Local Similarity   99.7%;  Pred. No. 4.8e-221;
  Matches  792;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     289 GCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCA 348
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   92092 GCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCA 92033

Qy     349 GCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACA 408
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   92032 GCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACA 91973

Qy     409 GGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCC 468
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   91972 GGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCC 91913

Qy     469 CCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATG 528
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   91912 CCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATG 91853

Qy     529 GGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGC 588
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   91852 GGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGC 91793

Qy     589 TGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGA 648
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   91792 TGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGA 91733

Qy     649 CACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGG 708
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   91732 CACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCCAGAAATAGGAAAGGCTAGGATTGG 91673

Qy     709 ATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGG 768
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  91672 ATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGG 91613

Qy    769 GAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACAT 828
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  91612 GAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACAT 91553

Qy    829 CCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCT 888
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  91552 CCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCT 91493

Qy    889 TCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGA 948
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  91492 TCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGA 91433

Qy    949 GAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCC 1008
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  91432 GAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCC 91373

Qy   1009 ATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACA 1068
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  91372 ATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACA 91313

Qy   1069 TTATTCACAAAAAA 1082
          |||||||||  ||||
Db  91312 TTATTCACAGAAAA 91299


RESULT   6
AC016708
LOCUS       AC016708              194304 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens BAC clone RP11-223M6 from 2, complete sequence.
ACCESSION   AC016708
VERSION     AC016708.11  GI:18093082
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194304)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 194304)
  AUTHORS   Yoakum,M., Harkins,C. and Haakenson,W.
  TITLE     The sequence of Homo sapiens BAC clone RP11-223M6
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 194304)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 194304)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jan 9, 2002 this sequence version replaced gi:13431120.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0223M06
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-172J1.  Actual start of
            this clone is at base position 1 of RP11-223M6; actual end is at
            base position 194304 of RP11-223M6.

            Data from AC068051 was used to finish the clone, AC016708.
            Polymorphisms have been identified between AC016708  and AC068051.
            An unresolved tandem repeat exists between 141687-142816.
            Approximately 600 bases are thought to be missing.
FEATURES             Location/Qualifiers
     source          1. .194304
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-223M6"
                     /clone_lib="RPCI-11"
     repeat_region   173. .204
                     /rpt_family="A-rich"
     repeat_region   294. .473
                     /rpt_family="MaLR"
     repeat_region   555. .716
                     /rpt_family="MaLR"
     repeat_region   1112. .1414
                     /rpt_family="MaLR"
     repeat_region   2301. .2326
                     /rpt_family="AT_rich"
     repeat_region   2695. .4442
                     /rpt_family="L1"
     misc_feature    3084. .3086
                     /note="similar to Homo sapiens EST AA765455 (NID:g2816693)
                     oa03a11.s1"
     misc_feature    4443. .4524
                     /note="similar to Homo sapiens EST BI770278
                     (NID:g15761856)"
     misc_feature    4445. .5222
                     /note="similar to Mus musculus EST BI156916
                     (NID:g14616917)"
     misc_feature    4445. .4524
                     /note="similar to Homo sapiens EST BG538360
                     (NID:g13530593)"
     misc_feature    4449. .4875
                     /note="similar to Mus musculus EST BE624006 (NID:g9904422)
                     uu46g10.y1"
     misc_feature    4450. .4524
                     /note="similar to Homo sapiens EST AI929700 (NID:g5665664)
                     au63f01.y1"
     misc_feature    4454. .4524
```

```
                /note="similar to Homo sapiens EST BI463649
                (NID:g15254305)"
misc_feature    4468. .4524
                /note="similar to Homo sapiens EST BI255301
                (NID:g14808547)"
misc_feature    4474. .5306
                /note="similar to Mus musculus EST BI661673
                (NID:g15575909)"
misc_feature    4578. .4870
                /note="similar to Homo sapiens EST BI255301
                (NID:g14808547)"
misc_feature    4578. .5003
                /note="similar to Homo sapiens EST BI770278
                (NID:g15761856)"
misc_feature    4578. .4917
                /note="similar to Homo sapiens EST AI929700 (NID:g5665664)
                au63f01.y1"
misc_feature    4578. .5222
                /note="similar to Homo sapiens EST BI463649
                (NID:g15254305)"
misc_feature    4578. .4875
                /note="similar to Homo sapiens EST BG538360
                (NID:g13530593)"
misc_feature    4636. .4917
                /note="similar to Homo sapiens EST BE830880
                (NID:g10263258)"
misc_feature    4636. .4821
                /note="similar to Homo sapiens EST AI658639 (NID:g4762209)
                tu06g05.x1"
misc_feature    4637. .5179
                /note="similar to Bos taurus EST AV599105 (NID:g9717850)"
misc_feature    4672. .5003
                /note="similar to Homo sapiens EST AA349861
                (NID:g2002181)"
misc_feature    4685. .4875
                /note="similar to Homo sapiens EST AW392207
                (NID:g6896866)"
misc_feature    4744
                /note="similar to Homo sapiens EST BG202037
                (NID:g13723724)"
misc_feature    4755. .5515
                /note="similar to Homo sapiens EST BG202037
                (NID:g13723724)"
misc_feature    4758. .5152
                /note="similar to Mus musculus EST BI664062
                (NID:g15578295)"
misc_feature    4858. .5232
                /note="similar to Homo sapiens EST H70078 (NID:g1040284)
                yr92b08.r1"
misc_feature    4892. .5499
                /note="similar to Homo sapiens EST BF997045
                (NID:g12403368)"
misc_feature    4894. .5062
                /note="similar to Homo sapiens EST BF992134
                (NID:g12398457)"
misc_feature    4922. .5507
                /note="similar to Homo sapiens EST BF997056
                (NID:g12403379)"
misc_feature    4943. .5500
                /note="similar to Homo sapiens EST BI015134
                (NID:g14419205)"
misc_feature    4987. .5719
                /note="similar to Homo sapiens EST BI856660
                (NID:g15997407)"
misc_feature    5101. .5520
                /note="similar to Homo sapiens EST AA765455 (NID:g2816693)
                oa03a11.s1"
misc_feature    5120. .5232
                /note="similar to Homo sapiens EST BF992134
                (NID:g12398457)"
misc_feature    5257. .5643
                /note="similar to Homo sapiens EST BF725771
                (NID:g12041682) bx19g05.y1"
misc_feature    5265. .5269
                /note="similar to Homo sapiens EST BF305685
                (NID:g11252669)"
misc_feature    5265. .5269
                /note="similar to Homo sapiens EST BF305826
                (NID:g11252835)"
misc_feature    5307. .5759
                /note="similar to Homo sapiens EST AI360592 (NID:g4112213)
                qx64d07.x1"
misc_feature    5391. .5923
                /note="similar to Homo sapiens EST BF305826

 Query Match             66.9%;  Score 738.2;  DB 9;  Length 194304;
 Best Local Similarity   84.6%;  Pred. No. 1.5e-205;
 Matches  918;  Conservative    0;  Mismatches  143;  Indels   24;  Gaps    7;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          |||||||||||| ||||||||||  |||||||||| |||||| ||||| ||||||| ||
Db    4468 GTGACAGGAGCCAAAGCAGCAGCATAGGTTGTCCCCATTTCCCTTCCCCTTTCCCTTGTC 4527

Qy      61 CGGTTGCCTTCCCGGGCCC-CTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAAC 119
           ||| | ||    |||||| | ||  | ||| || | | ||||  |||  |||||| |||
Db    4528 AGGTGGACTATATGGGCCCTCCTATGCCCCAGAGACTCAGTCCTACCACATCCCAGTAAC 4587

Qy     120 AAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAG 179
           |||||  |||||||||||| |||| |||||  | ||||||| | | | | |||||| ||||
Db    4588 AAGAAAGAGAGAACCCTGTGGAGAAGACCAACAAGGAGAAAC-GAAGATGCAGGATAAAG 4646

Qy     180 AAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAG 239
           ||||||||||  |||||||||  ||| ||| ||||||||||| |||||||| ||||||||
Db    4647 AAGGTATTCTCTCTGAGAGAGTTGAGGAGCCAAAGCTAAAGTCCAAATACCTAAGCCTAG 4706

Qy     240 GACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACT 299
           ||||| |||| ||||||||| |||||||||||||||| | |||||||||  ||| |||||
Db    4707 GACAAGAGCCCGGAGGCTCCAACTTCCTCATGAAGAGATTACAGAAAGGGAAAAGGTACT 4766

Qy     300 TTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTG 359
           ||| ||||||||||||||||||||||||||||||||||||||||| ||||||||| |||||||
Db    4767 TTGTCTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAGTAAGCAGCTTCCAAGTG 4826

Qy     360 CAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCC 419
           ||| |||| | |||||||||||||||||||| |||||                ||||||||
Db    4827 CAGAACCAAAGAAGAACCTGGTGACTGGTGACCACAT---------------TCTGCCCC 4871

Qy     420 AGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTC 479
           | |||||||||| | |  |||||||||||||||| ||||||||   ||  ||||||||||
Db    4872 ACAGAAAGTCCTTGGTTATCACCAGCAAGCTTGCAGGGTAACCCAGGCTGCCCTCTCCTC 4931

Qy     480 CCCTTCCTCA-ACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGT 538
           |   ||||||  |||||||| ||||| ||||||||||||||||||||||||| ||||||||
Db    4932 CTACTCCTCACCCCACTGGATGTTTACATATTATAGGCAGGGATGAAATGGACACCTAGT 4991

Qy     539 CAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAG 598
           |||||||       || | || || |||||||| ||||||| || | |||| ||||||||
Db    4992 CAGATCT----CAGTTACCAGGCAAAAATGACTATGATTCTCCTAGTGGCTTCTGAGAAG 5047

Qy     599 GTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACA 658
           |||||||| ||||||  ||| |||| |||  |||| ||| |||| ||||| ||| | |||
Db    5048 GTAATGTAAGTTGAAGGGGGACTCTGAGTCCATTTCTTTGGTTAAATTGAGACTCCTACA 5107

Qy     659 CACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGC 718
             |||||| || |||||||||||||| |||||||||||||| ||||||||||||||||||
Db    5108 TCCTCCCTATAATCCAGGTAGGGCCCAGAAATAGGAAAGACTAGGATTGGATAATGCTGC 5167

Qy     719 AAATGCTTT-TTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATG 777
           |||||||| | || ||||||||||||||||||||||||||||||||||  ||||||
Db    5168 AAATGCTTTCTCTTTTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGAGGAGAATA 5227

Qy     778 TCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCT 837
           ||||| |||||||| |||||||||||||||| |  |||||||||||||| || |||
Db    5228 TCCCATAATTGATTAAGCTGAGCCTTGGGAATAATACGGCAGGTTTAACATCCAAAAGCT 5287

Qy     838 AACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGT 897
```

```
Db    5288 AACCTGACATAGTTGGGAAAGGTAGATTAAATAAGACATGTTTT-TGTGCTTCTAAGTGT 5346

Qy     898 TCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTA 957

Db    5347 TTTGTCCCTTATGCTGCTATTGCTTCATGTTTCCATTACGGCAGGTTTAGAAAATCCTTA 5406

Qy     958 AAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAG 1017

Db    5407 CAAAGAAAGACAGACTTGCTTGCCTAAAACTACAATGTCCACTTAGCCCCCTTTACTTAG 5466

Qy    1018 TATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACATTATTCACA 1077

Db    5467 TATCTCTTGCAGTTTGCCCTGGCTCTCAAATAATGTAAAGATTGATGAACATTACTCACA 5526

Qy    1078 AAAAA 1082

Db    5527 GAAAA 5531


RESULT   7
HSALPEND
LOCUS       HSALPEND                749 bp    mRNA    linear   PRI 22-SEP-1998
DEFINITION  Homo sapiens mRNA for alpha endosulfine.
ACCESSION   X99906
VERSION     X99906.1  GI:2764973
KEYWORDS    alpha endosulfine; sulfonylurea receptor.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 749)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-1996) D. Bataille, Inserm U 376, Endocrinologie
            des Peptides, 371, Rue du Doyen Gaston Giraud, Montpellier Cedex
            05, 34294, FRANCE
REFERENCE   2  (bases 1 to 749)
  AUTHORS   Heron,L., Virsolvy,A., Peyrollier,K., Gribble,F.M., Le Cam,A.,
            Ashcroft,F.M. and Bataille,D.
  TITLE     Human alpha-endosulfine, a possible regulator of
            sulfonylurea-sensitive KATP channel: molecular cloning, expression
            and biological properties
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (14), 8387-8391 (1998)
  MEDLINE   98318667
FEATURES             Location/Qualifiers
     source          1. .749
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="brain"
                     /dev_stage="adult"
     CDS             126. .491
                     /function="endogenous lignad for sulfonylurea receptor"
                     /codon_start=1
                     /product="alpha endosulfine"
                     /protein_id="CAA68180.1"
                     /db_xref="GI:2764974"
                     /db_xref="SPTREMBL:O43768"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAGGQVE"
BASE COUNT      184 a    224 c    200 g    141 t
ORIGIN


  Query Match             41.3%;  Score 456;  DB 9;  Length 749;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-123;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60

Db      20 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 79

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120

Db      80 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 139

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180

Db     140 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 199

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240

Db     200 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 259

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300

Db     260 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 319

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360

Db     320 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 379

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420

Db     380 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 439

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456

Db     440 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 475


RESULT   8
AF157509
LOCUS       AF157509                1199 bp    mRNA    linear   PRI 01-JUL-2000
DEFINITION  Homo sapiens alpha endosulfine mRNA, complete cds.
ACCESSION   AF157509
VERSION     AF157509.1  GI:8886008
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1199)
  AUTHORS   Scott,V.E.S., Roch,J.-M., Davis-Taber,R.A., Molinari,E.J.,
            Whiteaker,K.L., Gopalakrishnan,M., Idler,K. and Sullivan,J.P.
  TITLE     Cloning and molecular characterization of two isoforms of human
            endosulfine
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1199)
  AUTHORS   Scott,V.E.S., Roch,J.-M., Whiteaker,K.L. and Sullivan,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-1999) D47C, Abbott Laboratories, AP10 L08, 100
            Abbott Park, Abbott Park, IL 60064, USA
FEATURES             Location/Qualifiers
     source          1. .1199
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             107. .472
                     /codon_start=1
                     /product="alpha endosulfine"
                     /protein_id="AAF80340.1"
                     /db_xref="GI:8886009"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAGGQVE"
BASE COUNT      335 a    288 c    333 g    243 t
ORIGIN


  Query Match             41.3%;  Score 456;  DB 9;  Length 1199;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-123;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
```

```
Db      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60

Qy     61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120

Qy    121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180

Qy    181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240

Qy    241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300

Qy    301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360

Qy    361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420

Qy    421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
          |||||||||||||||||||||||||||||||||||
Db    421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
```

```
RESULT  9
AK001981
LOCUS       AK001981               1208 bp    mRNA    linear   PRI 22-FEB-2000
DEFINITION  Homo sapiens cDNA FLJ11119 fis, clone PLACE1006040, highly similar
            to Homo sapiens mRNA for alpha endosulfine.
ACCESSION   AK001981
VERSION     AK001981.1  GI:7023587
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens placenta cDNA to mRNA, clone_lib:PLACE1
            clone:PLACE1006040.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K.,
            Tanai,H., Kimata,M., Watanabe,M., Hiraoka,S., Ishii,S., Kawai,Y.,
            Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahari,K.,
            Masuho,Y. and Kanehori,K.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 1208)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao
            Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
            Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1. .1208
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PLACE1006040"
                     /clone_lib="PLACE1"
                     /tissue_type="placenta"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      326 a    295 c    337 g    250 t
ORIGIN
```

```
  Query Match             41.3%;  Score 456;  DB 9;  Length 1208;
  Best Local Similarity   100.0%;  Pred. No. 6e-123;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 80

Qy     61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 140

Qy    121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 200

Qy    181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 260

Qy    241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 320

Qy    301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 380

Qy    361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 440

Qy    421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
          |||||||||||||||||||||||||||||||||||
Db    441 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 476
```

```
RESULT  10
HSAJ10966
LOCUS       HSAJ10966              1788 bp    DNA     linear   PRI 07-JAN-2000
DEFINITION  Homo sapiens ENSA gene.
ACCESSION   AJ010966
VERSION     AJ010966.1  GI:6688263
KEYWORDS    alpha endosulfine; ENSA gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1788)
  AUTHORS   Heron,L., Virsolvy,A., Peyrollier,K., Gribble,F.M., Le Cam,A.,
            Ashcroft,F.M. and Bataille,D.
  TITLE     Human alpha-endosulfine, a possible regulator of
            sulfonylurea-sensitive KATP channel: molecular cloning, expression
            and biological properties
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (14), 8387-8391 (1998)
  MEDLINE   98318667
REFERENCE   2  (bases 1 to 1788)
  AUTHORS   Bataille,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1788)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-1998) Bataille D., INSERM, U 376, CHU
            Arnaud-de-Villeneuve, Montpellier, 34295, FRANCE
  REMARK    revised by author
FEATURES             Location/Qualifiers
     source          1. .1788
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
     gene            757. .1122
                     /gene="ENSA"
     CDS             757. .1122
                     /gene="ENSA"
                     /function="putative regulator of K-ATP channels"
                     /codon_start=1
                     /evidence=experimental
                     /product="alpha endosulfine"
                     /protein_id="CAB65125.1"
                     /db_xref="GI:6688264"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAGGQVE"
BASE COUNT      532 a    388 c    411 g    457 t
ORIGIN


  Query Match           41.2%;  Score 454.4;  DB 9;  Length 1788;
  Best Local Similarity   99.8%;  Pred. No. 1.9e-122;
  Matches  455;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    651 GTGATAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 710

Qy     61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    711 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 770

Qy    121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    771 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 830

Qy    181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    831 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 890

Qy    241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    891 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 950

Qy    301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    951 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 1010

Qy    361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1011 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 1070

Qy    421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
          |||||||||||||||||||||||||||||||||||
Db   1071 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 1106


RESULT  11
BC000436
LOCUS       BC000436               1211 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, endosulfine alpha, clone MGC:8394 IMAGE:2820528,
            mRNA, complete cds.
ACCESSION   BC000436
VERSION     BC000436.1  GI:12653334
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1211)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site:       http://www.nisc.nih.gov/
            Contact:        nisc_mgc@nhgri.nih.gov
            Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R.,
            Lim,M., Maduro,Q.L., Masiello,C., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Snyder,B., Stantripop,S., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 1 Row: d Column: 1
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 8886008.
FEATURES              Location/Qualifiers
     source           1. .1211
                      /organism="Homo sapiens"
                      /db_xref="LocusID:2029"
                      /db_xref="taxon:9606"
                      /clone="MGC:8394 IMAGE:2820528"
                      /tissue_type="Lung, small cell carcinoma"
                      /clone_lib="NIH_MGC_7"
                      /lab_host="DH10B-R"
                      /note="Vector: pOTB7"
     CDS              110. .475
                      /codon_start=1
                      /product="endosulfine alpha"
                      /protein_id="AAH00436.1"
                      /db_xref="GI:12653335"
                      /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL
                      GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD
                      LPQRKSSLVTSKLAGGQVE"
BASE COUNT      340 a    291 c    337 g    243 t
ORIGIN


  Query Match           40.9%;  Score 450.8;  DB 9;  Length 1211;
  Best Local Similarity   99.6%;  Pred. No. 2e-121;
  Matches  452;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 GACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCG 62
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6 GAGGGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCG 65

Qy     63 GTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAG 122
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66 GTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAG 125

Qy    123 AAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAG 182
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    126 AAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAG 185

Qy    183 GTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGAC 242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    186 GTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGAC 245

Qy    243 AAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTG 302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    246 AAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTG 305

Qy    303 ACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAG 362
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    306 ACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAG 365

Qy    363 GACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGA 422
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    366 GACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGA 425

Qy    423 GAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
          |||||||||||||||||||||||||||||||||
Db    426 GAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 459


RESULT  12
AC091596
LOCUS       AC091596               69732 bp    DNA     linear   HTG 28-NOV-2001
DEFINITION  Homo sapiens chromosome 17 clone RP11-474K4 map 17, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC091596
VERSION     AC091596.2  GI:17047214
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 69732)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-474K4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 69732)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 22, 2001 this sequence version replaced gi:13959213.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L13230
                Center clone name: 474_K_4
            --------------.
            * NOTE: This record contains 87 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      703: contig of 703 bp in length
            *      704 803: gap of      100 bp
            *      804     1494: contig of 691 bp in length
            *     1495 1594: gap of      100 bp
            *     1595     2369: contig of 775 bp in length
            *     2370 2469: gap of      100 bp
            *     2470     3158: contig of 689 bp in length
            *     3159 3258: gap of      100 bp
            *     3259     3977: contig of 719 bp in length
            *     3978 4077: gap of      100 bp
            *     4078     4743: contig of 666 bp in length
            *     4744 4843: gap of      100 bp
            *     4844     5563: contig of 720 bp in length
            *     5564 5663: gap of      100 bp
            *     5664     6345: contig of 682 bp in length
            *     6346 6445: gap of      100 bp
            *     6446     7156: contig of 711 bp in length
            *     7157 7256: gap of      100 bp
            *     7257     7963: contig of 707 bp in length
            *     7964 8063: gap of      100 bp
            *     8064     8750: contig of 687 bp in length
            *     8751 8850: gap of      100 bp
            *     8851     9519: contig of 669 bp in length
            *     9520 9619: gap of      100 bp
            *     9620    10352: contig of 733 bp in length
            *    10353 10452: gap of      100 bp
            *    10453    11154: contig of 702 bp in length
            *    11155 11254: gap of      100 bp
            *    11255    11937: contig of 683 bp in length
            *    11938 12037: gap of      100 bp
            *    12038    12730: contig of 693 bp in length
            *    12731 12830: gap of      100 bp
            *    12831    13527: contig of 697 bp in length
            *    13528 13627: gap of      100 bp
            *    13628    14348: contig of 721 bp in length
            *    14349 14448: gap of      100 bp
            *    14449    15174: contig of 726 bp in length
            *    15175 15274: gap of      100 bp
            *    15275    16001: contig of 727 bp in length
            *    16002 16101: gap of      100 bp
            *    16102    16799: contig of 698 bp in length
            *    16800 16899: gap of      100 bp
            *    16900    17580: contig of 681 bp in length
            *    17581 17680: gap of      100 bp
            *    17681    18359: contig of 679 bp in length
            *    18360 18459: gap of      100 bp
            *    18460    19148: contig of 689 bp in length
            *    19149 19248: gap of      100 bp
            *    19249    19935: contig of 687 bp in length
            *    19936 20035: gap of      100 bp
            *    20036    20718: contig of 683 bp in length
            *    20719 20818: gap of      100 bp
            *    20819    21535: contig of 717 bp in length
            *    21536 21635: gap of      100 bp
            *    21636    22353: contig of 718 bp in length
            *    22354 22453: gap of      100 bp
            *    22454    23163: contig of 710 bp in length
            *    23164 23263: gap of      100 bp
            *    23264    23967: contig of 704 bp in length
            *    23968 24067: gap of      100 bp
            *    24068    24768: contig of 701 bp in length
            *    24769 24868: gap of      100 bp
            *    24869    25545: contig of 677 bp in length
            *    25546 25645: gap of      100 bp
            *    25646    26366: contig of 721 bp in length
            *    26367 26466: gap of      100 bp
            *    26467    27158: contig of 692 bp in length
            *    27159 27258: gap of      100 bp
            *    27259    27958: contig of 700 bp in length
            *    27959 28058: gap of      100 bp
```

```
*    28059    28760: contig of 702 bp in length
*    28761 28860: gap of      100 bp
*    28861    29590: contig of 730 bp in length
*    29591 29690: gap of      100 bp
*    29691    30413: contig of 723 bp in length
*    30414 30513: gap of      100 bp
*    30514    31242: contig of 729 bp in length
*    31243 31342: gap of      100 bp
*    31343    32073: contig of 731 bp in length
*    32074 32173: gap of      100 bp
*    32174    32876: contig of 703 bp in length
*    32877 32976: gap of      100 bp
*    32977    33670: contig of 694 bp in length
*    33671 33770: gap of      100 bp
*    33771    34469: contig of 699 bp in length
*    34470 34569: gap of      100 bp
*    34570    35271: contig of 702 bp in length
*    35272 35371: gap of      100 bp
*    35372    36066: contig of 695 bp in length
*    36067 36166: gap of      100 bp
*    36167    36867: contig of 701 bp in length
*    36868 36967: gap of      100 bp
*    36968    37670: contig of 703 bp in length
*    37671 37770: gap of      100 bp
*    37771    38497: contig of 727 bp in length
*    38498 38597: gap of      100 bp
*    38598    39313: contig of 716 bp in length
*    39314 39413: gap of      100 bp
*    39414    40135: contig of 722 bp in length
*    40136 40235: gap of      100 bp
*    40236    40925: contig of 690 bp in length
*    40926 41025: gap of      100 bp
*    41026    41728: contig of 703 bp in length
*    41729 41828: gap of      100 bp
*    41829    42518: contig of 690 bp in length
*    42519 42618: gap of      100 bp
*    42619    43314: contig of 696 bp in length
*    43315 43414: gap of      100 bp
*    43415    44112: contig of 698 bp in length
*    44113 44212: gap of      100 bp
*    44213    44908: contig of 696 bp in length
*    44909 45008: gap of      100 bp
*    45009    45707: contig of 699 bp in length
*    45708 45807: gap of      100 bp
*    45808    46528: contig of 721 bp in length
*    46529 46628: gap of      100 bp
*    46629    47355: contig of 727 bp in length
*    47356 47455: gap of      100 bp
*    47456    48179: contig of 724 bp in length
*    48180 48279: gap of      100 bp
*    48280    48992: contig of 713 bp in length
*    48993 49092: gap of      100 bp
*    49093    49803: contig of 711 bp in length
*    49804 49903: gap of      100 bp
*    49904    50605: contig of 702 bp in length
*    50606 50705: gap of      100 bp
*    50706    51397: contig of 692 bp in length
*    51398 51497: gap of      100 bp
*    51498    52196: contig of 699 bp in length
*    52197 52296: gap of      100 bp
*    52297    52999: contig of 703 bp in length
*    53000 53099: gap of      100 bp
*    53100    53787: contig of 688 bp in length
*    53788 53887: gap of      100 bp
*    53888    54545: contig of 658 bp in length
*    54546 54645: gap of      100 bp
*    54646    55332: contig of 687 bp in length

Query Match             40.9%;  Score 450.6;  DB 2;  Length 69732;
Best Local Similarity   73.0%;  Pred. No. 6.2e-121;
Matches  590;  Conservative    0;  Mismatches  195;  Indels   23;  Gaps    6;

Qy    110 TCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACG 169
          | |||| ||||||||  |||||||||| |||| |||||  | ||||||| | | | | |
Db  24087 TTCCAGTAACAAGAAAGAGAGAACCCTGTGGAGAAGACCAACAAGGAGAAAC-GAAGATG 24145

Qy    170 CAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATAC 229
          |||||  |||||||||||||||  |||||||| ||| ||| |||||||||| ||||||||
Db  24146 CAGGATAAAGAAGGTATTCTCTCTGAGAGAGTTGAGGAGCCAAAGCTAAAGTCCAAATAC 24205

Qy    230 CCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGG 289
          | |||||||||||||  |||| |||||||| |||||||||||||||||| | |||||||||
Db  24206 CTAAGCCTAGGACAAGAGCCCGGAGGCTCCAACTTCCTCATGAAGAGATTACAGAAAGGG 24265

Qy    290 CAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAG 349
            ||| |||||||| ||||||||||||||||||||||||||||||||||||||| |||||||
Db  24266 AAAAGGTACTTTGTCTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAGTAAGCAG 24325

Qy    350 CTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAG 409
          || |||||||||| |||| | ||||||||||||||||||| |||||
Db  24326 CTTCCAAGTGCAGAACCAAAGAAGAACCTGGTGACTGGTGACCACAT------------ 24372

Qy    410 GATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCC 469
            |||||||| |||||||||| | |  |||||||||||||||| ||||||||   ||
Db  24373 --TCTGCCCCACAGAAAGTCCTTGGTTATCACCAGCAAGCTTGCAGGGTAACCCAGGCTG 24430

Qy    470 CCCTCTCCTCCCCTTCCTCA-ACCACTGGACGTTTATATATTATAGGCAGGGATGAAATG 528
          ||||||||||   ||||||  |||||||| ||||| ||||||||||||||||||||||
Db  24431 CCCTCTCCTCCTACTCCTCACCCCACTGGATGTTTACATATTATAGGCAGGGATGAAATG 24490

Qy    529 GGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGC 588
          | ||||||||||||||        || | || || |||||||| ||||||| || | |||
Db  24491 GACACCTAGTCAGATCT----CAGTTACCAGGCAAAAATGACTATGATTCTCCTAGTGGC 24546

Qy    589 TGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGA 648
          | |||||||||||||||| ||||||  ||| |||| |||  |||| ||| |||| |||||
Db  24547 TTCTGAGAAGGTAATGTAAGTTGAAGGGGGACTCTGAGTCCATTTCTTTGGTTAAATTGA 24606

Qy    649 CACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGG 708
          ||| | |||  |||||| || |||||||||||| |||||||||||| ||||||||||
Db  24607 GACTCCTACATCCTCCCTATAATCCAGGTAGGGCCCAGAAATAGGAAAGACTAGGATTGG 24666

Qy    709 ATAATGCTGCAAATGCTTT-TTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTG 767
          |||||||||||||||||| | || ||||||||||||||||||||||||||||||||||||
Db  24667 ATAATGCTGCAAATGCTTTCTCTTTTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTG 24726

Qy    768 GGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACA 827
          |  ||| || |||| | |||||||| |||||||||| |||||
Db  24727 GAGGAG-ATATCCCCATATTGATTAAGCTGAGCCTTGNGAATANNNNNNNNNNNNNNNNN 24785

Qy    828 TCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGC 887

Db  24786 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 24845

Qy    888 TTCTAAGTGTTCTGTCCCTTAGGCTGCT 915
                              || |
Db  24846 NNNNNNNNNNNNNNNNNNNNNNNGTGGT 24873


RESULT  13
BC004461
LOCUS       BC004461                1202 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, endosulfine alpha, clone MGC:4319 IMAGE:2820528,
            mRNA, complete cds.
ACCESSION   BC004461
VERSION     BC004461.1  GI:13325294
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1202)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
```

```
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 9 Row: a Column: 17
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 8886008.
FEATURES              Location/Qualifiers
     source           1. .1202
                      /organism="Homo sapiens"
                      /db_xref="LocusID:2029"
                      /db_xref="taxon:9606"
                      /clone="MGC:4319 IMAGE:2820528"
                      /tissue_type="Lung, small cell carcinoma"
                      /clone_lib="NIH_MGC_7"
                      /lab_host="DH10B-R"
                      /note="Vector: pOTB7"
     CDS              101. .466
                      /codon_start=1
                      /product="endosulfine alpha"
                      /protein_id="AAH04461.1"
                      /db_xref="GI:13325295"
                      /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL
                      GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD
                      LPQRKSSLVTSKLAGGQVE"
BASE COUNT      338 a    289 c    332 g    243 t
ORIGIN


  Query Match             40.8%;  Score 450;  DB 9;  Length 1202;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-121;
  Matches  450;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 GGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTG 66
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTG 60

Qy      67 CCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGA 126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGA 120

Qy     127 AGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTAT 186
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTAT 180

Qy     187 TCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAA 246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAA 240

Qy     247 GCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTC 306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTC 300

Qy     307 AGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACC 366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACC 360

Qy     367 AGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAA 426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAA 420

Qy     427 GTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
          |||||||||||||||||||||||||||||
Db     421 GTCCTCGCTCGTCACCAGCAAGCTTGCGGG 450


RESULT  14
AR144151
LOCUS       AR144151                669 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 16 from patent US 6210883.
ACCESSION   AR144151
VERSION     AR144151.1  GI:15106018
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   Reed,S.G. and Wang,T.Tong.
  TITLE     Compounds and methods for diagnosis of lung cancer
  JOURNAL   Patent: US 6210883-A 16 03-APR-2001;
FEATURES              Location/Qualifiers
     source           1. .669
                      /organism="unknown"
BASE COUNT      158 a    208 c    173 g    119 t     11 others
ORIGIN


  Query Match             39.1%;  Score 431.4;  DB 6;  Length 669;
  Best Local Similarity   99.1%;  Pred. No. 9.1e-116;
  Matches  443;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy      10 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 69
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 61

Qy      70 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 129
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 121

Qy     130 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 189
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 181

Qy     190 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 249
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 241

Qy     250 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 309
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     242 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCNGG 301

Qy     310 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 369
          |||||||||||||||||||||| |||||||||||||||||||||||||||| |||||||
Db     302 AGACTACAACATGGCCAAAGCCAACATGAAGAATAAGCAGCTGCCAAGTGCANGACCAGA 361

Qy     370 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 429
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     362 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTG-CCCAGAGAAAGTC 420

Qy     430 CTCGCTCGTCACCAGCAAGCTTGCGGG 456
          ||||||||||||||||||||||||||||
Db     421 CTCGCTCGTCACCAGCAAGCTTGCGGG 447


RESULT  15
AR176338
LOCUS       AR176338                669 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 16 from patent US 6312695.
```

```
ACCESSION   AR176338
VERSION     AR176338.1  GI:17918693
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   Reed,S.G. and Wang,T.Tong.
  TITLE     Compounds and methods for therapy of lung cancer
  JOURNAL   Patent: US 6312695-A 16 06-NOV-2001;
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="unknown"
BASE COUNT      158 a    208 c    173 g    119 t     11 others
ORIGIN
```

Query Match 39.1%; Score 431.4; DB 6; Length 669;
Best Local Similarity 99.1%; Pred. No. 9.1e-116;
Matches 443; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```
Qy      10 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 61

Qy      70 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 121

Qy     130 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 181

Qy     190 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 241

Qy     250 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     242 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCNGG 301

Qy     310 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 369
           |||||||||||||||||||||| |||||||||||||||||||||||||||| |||||||
Db     302 AGACTACAACATGGCCAAAGCCAACATGAAGAATAAGCAGCTGCCAAGTGCANGACCAGA 361

Qy     370 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 429
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     362 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTG-CCCAGAGAAAGTC 420

Qy     430 CTCGCTCGTCACCAGCAAGCTTGCGGG 456
           ||||||||||||||||||||||||||
Db     421 CTCGCTCGTCACCAGCAAGCTTGCGGG 447
```

```
RESULT  16
AC023294
LOCUS       AC023294               98113 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-17L18, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC023294
VERSION     AC023294.2  GI:9123869
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 98113)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-17L18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 98113)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6960434.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3656
                Center clone name: 17_L_18
            --------------.
            * NOTE: This record contains 90 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      996: contig of 996 bp in length
            *      997 1096: gap of      100 bp
            *     1097     2083: contig of 987 bp in length
            *     2084 2183: gap of      100 bp
            *     2184     3189: contig of 1006 bp in length
            *     3190 3289: gap of      100 bp
            *     3290     4298: contig of 1009 bp in length
            *     4299 4398: gap of      100 bp
            *     4399     5381: contig of 983 bp in length
            *     5382 5481: gap of      100 bp
            *     5482     6423: contig of 942 bp in length
            *     6424 6523: gap of      100 bp
            *     6524     7541: contig of 1018 bp in length
            *     7542 7641: gap of      100 bp
            *     7642     8604: contig of 963 bp in length
            *     8605 8704: gap of      100 bp
            *     8705     9669: contig of 965 bp in length
            *     9670 9769: gap of      100 bp
            *     9770    10781: contig of 1012 bp in length
            *    10782 10881: gap of      100 bp
            *    10882    11829: contig of 948 bp in length
            *    11830 11929: gap of      100 bp
            *    11930    12922: contig of 993 bp in length
            *    12923 13022: gap of      100 bp
            *    13023    14013: contig of 991 bp in length
            *    14014 14113: gap of      100 bp
            *    14114    15128: contig of 1015 bp in length
            *    15129 15228: gap of      100 bp
```

```
*    15229    16225: contig of 997 bp in length
*    16226 16325: gap of      100 bp
*    16326    17325: contig of 1000 bp in length
*    17326 17425: gap of      100 bp
*    17426    18419: contig of 994 bp in length
*    18420 18519: gap of      100 bp
*    18520    19517: contig of 998 bp in length
*    19518 19617: gap of      100 bp
*    19618    20592: contig of 975 bp in length
*    20593 20692: gap of      100 bp
*    20693    21693: contig of 1001 bp in length
*    21694 21793: gap of      100 bp
*    21794    22784: contig of 991 bp in length
*    22785 22884: gap of      100 bp
*    22885    23897: contig of 1013 bp in length
*    23898 23997: gap of      100 bp
*    23998    24980: contig of 983 bp in length
*    24981 25080: gap of      100 bp
*    25081    26067: contig of 987 bp in length
*    26068 26167: gap of      100 bp
*    26168    27142: contig of 975 bp in length
*    27143 27242: gap of      100 bp
*    27243    28264: contig of 1022 bp in length
*    28265 28364: gap of      100 bp
*    28365    29384: contig of 1020 bp in length
*    29385 29484: gap of      100 bp
*    29485    30487: contig of 1003 bp in length
*    30488 30587: gap of      100 bp
*    30588    31595: contig of 1008 bp in length
*    31596 31695: gap of      100 bp
*    31696    32706: contig of 1011 bp in length
*    32707 32806: gap of      100 bp
*    32807    33747: contig of 941 bp in length
*    33748 33847: gap of      100 bp
*    33848    34860: contig of 1013 bp in length
*    34861 34960: gap of      100 bp
*    34961    35923: contig of 963 bp in length
*    35924 36023: gap of      100 bp
*    36024    36997: contig of 974 bp in length
*    36998 37097: gap of      100 bp
*    37098    38112: contig of 1015 bp in length
*    38113 38212: gap of      100 bp
*    38213    39249: contig of 1037 bp in length
*    39250 39349: gap of      100 bp
*    39350    40364: contig of 1015 bp in length
*    40365 40464: gap of      100 bp
*    40465    41450: contig of 986 bp in length
*    41451 41550: gap of      100 bp
*    41551    42542: contig of 992 bp in length
*    42543 42642: gap of      100 bp
*    42643    43654: contig of 1012 bp in length
*    43655 43754: gap of      100 bp
*    43755    44761: contig of 1007 bp in length
*    44762 44861: gap of      100 bp
*    44862    45833: contig of 972 bp in length
*    45834 45933: gap of      100 bp
*    45934    46886: contig of 953 bp in length
*    46887 46986: gap of      100 bp
*    46987    47948: contig of 962 bp in length
*    47949 48048: gap of      100 bp
*    48049    49034: contig of 986 bp in length
*    49035 49134: gap of      100 bp
*    49135    50128: contig of 994 bp in length
*    50129 50228: gap of      100 bp
*    50229    51198: contig of 970 bp in length
*    51199 51298: gap of      100 bp
*    51299    52314: contig of 1016 bp in length
*    52315 52414: gap of      100 bp
*    52415    53389: contig of 975 bp in length
*    53390 53489: gap of      100 bp
*    53490    54509: contig of 1020 bp in length
*    54510 54609: gap of      100 bp
*    54610    55578: contig of 969 bp in length
*    55579 55678: gap of      100 bp
*    55679    56672: contig of 994 bp in length
*    56673 56772: gap of      100 bp
*    56773    57768: contig of 996 bp in length
*    57769 57868: gap of      100 bp
*    57869    58834: contig of 966 bp in length
*    58835 58934: gap of      100 bp
*    58935    59929: contig of 995 bp in length
*    59930 60029: gap of      100 bp
*    60030    61038: contig of 1009 bp in length
*    61039 61138: gap of      100 bp
*    61139    62129: contig of 991 bp in length
*    62130 62229: gap of      100 bp
*    62230    63225: contig of 996 bp in length
*    63226 63325: gap of      100 bp
*    63326    64369: contig of 1044 bp in length
*    64370 64469: gap of      100 bp
*    64470    65464: contig of 995 bp in length
*    65465 65564: gap of      100 bp
*    65565    66575: contig of 1011 bp in length
*    66576 66675: gap of      100 bp
*    66676    67664: contig of 989 bp in length
*    67665 67764: gap of      100 bp
*    67765    68797: contig of 1033 bp in length
*    68798 68897: gap of      100 bp
*    68898    69897: contig of 1000 bp in length
*    69898 69997: gap of      100 bp
*    69998    71002: contig of 1005 bp in length
*    71003 71102: gap of      100 bp
*    71103    72121: contig of 1019 bp in length
*    72122 72221: gap of      100 bp
*    72222    73257: contig of 1036 bp in length
*    73258 73357: gap of      100 bp
*    73358    74351: contig of 994 bp in length
*    74352 74451: gap of      100 bp
*    74452    75462: contig of 1011 bp in length
*    75463 75562: gap of      100 bp

 Query Match             31.0%;  Score 342.2;  DB 2;  Length 98113;
 Best Local Similarity   81.7%;  Pred. No. 5.8e-89;
 Matches  447;  Conservative    0;  Mismatches   82;  Indels   18;  Gaps    4;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           |||||||||||| |||||||||| |||||||||| |||||| ||||| ||||||| ||
Db   89944 GTGACAGGAGCCAAAGCAGCAGCATAGGTTGTCCCCATTTCCCTTCCCCTTTCCCTTGTC 90003

Qy      61 CGGTTGCCTTCCCGGGCCC-CTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAAC 119
            ||| | ||    |||||| | ||  | ||| || | | |||| ||| |||||| |||
Db   90004 AGGTGGACTATATGGGCCCTCCTATGCCCCAGAGACTCAGTCCTACCACATCCCAGTAAC 90063

Qy     120 AAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAG 179
           |||||  |||||||||||| |||| |||||  | ||||||| | | | | |||||| ||||
Db   90064 AAGAAAGAGAGAACCCTGTGGAGAAGACCAACAAGGAGAAAC-GAAGATGCAGGATAAAG 90122

Qy     180 AAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAG 239
           ||||||||||  ||||||||| ||| ||| ||||||||||| |||||||||| ||||||||
Db   90123 AAGGTATTCTCTCTGAGAGAGTTGAGGAGCCAAAGCTAAAGTCCAAATACCTAAGCCTAG 90182

Qy     240 GACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACT 299
           ||||| |||| ||||||||| ||||||||||||||||| | |||||||||| ||| |||||
Db   90183 GACAAGAGCCCGGAGGCTCCAACTTCCTCATGAAGAGATTACAGAAAGGGAAAAGGTACT 90242

Qy     300 TTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTG 359
           ||| ||||||||||||||||||||||||||||||||||||| ||||||||| |||||||
Db   90243 TTGTCTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAGTAAGCAGCTTCCAAGTG 90302

Qy     360 CAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCC 419
           ||| |||| | |||||||||||||||||| ||||| |||||||||
Db   90303 CAGAACCAAAGAAGAACCTGGTGACTGGTGACCACAT----------TCTGCCCC 90347

Qy     420 AGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTC 479
           | ||||||| || | |  |||||||||||||||| |||||||| || |||||| ||
Db   90348 ACAGAAAGTNCTTGGTTATCACCAGCAAGCTTGCAGGGTAACCCAGGCTGCCCTCTNCTN 90407
```

```
Qy    480 CCCTTCCTCA-ACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGT 538
          |   | ||||  |||||||| ||||| |||||||||||||||||||||||| ||||||||
Db  90408 CTACTNCTCACCCCACTGGATGTTTACATATTATAGGCAGGGATGAAATGGACACCTAGT 90467

Qy    539 CAGATCT 545
          |||||||
Db  90468 CAGATCT 90474


RESULT  17
SSA005987
LOCUS       SSA005987               366 bp    mRNA    linear   MAM 07-JAN-1999
DEFINITION  Sus scrofa alpha-endosulfine (complete coding sequence).
ACCESSION   AJ005987
VERSION     AJ005987.1  GI:4138556
KEYWORDS    alpha-endosulfine.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Bataille D., U 376 'Endocrinologie des
            Peptides et regulation genique', INSERM, CHU Arnaud-de-Villeneuve,
            34295, FRANCE
REFERENCE   2  (bases 1 to 366)
  AUTHORS   Virsolvy-Vergine,A., Salazar,G., Sillard,R., Denoroy,L., Mutt,V.
            and Bataille,D.
  TITLE     Endosulfine, endogenous ligand for the sulphonylurea receptor:
            isolation from porcine brain and partial structural determination
            of the alpha form
  JOURNAL   Diabetologia 39 (2), 135-141 (1996)
  MEDLINE   96218590
REFERENCE   3  (bases 1 to 366)
  AUTHORS   Peyrollier,K., Heron,L., Virsolvy-Vergine,A., Le Cam,A. and
            Bataille,D.
  TITLE     Alpha endosulfine is a novel molecule, structurally related to a
            family of phosphoproteins
  JOURNAL   Biochem. Biophys. Res. Commun. 223 (3), 583-586 (1996)
  MEDLINE   96268722
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /tissue_type="brain"
     CDS             1. .366
                     /function="endogenous regulator of K-ATP channels"
                     /codon_start=1
                     /evidence=experimental
                     /product="alpha-endosulfine"
                     /protein_id="CAA06801.1"
                     /db_xref="GI:4138557"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPEKAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAGGQVE"
BASE COUNT      129 a     91 c     96 g     50 t
ORIGIN


  Query Match           30.3%;  Score 334;  DB 4;  Length 366;
  Best Local Similarity   97.1%;  Pred. No. 3.9e-87;
  Matches  340;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy    107 ATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGAC 166
          ||||||||||||||||||||| ||||||||||||| ||||||||||||||||||||
Db      1 ATGTCCCAGAAACAAGAAGAAGAAAACCCTGCGGAGGAAACCGGCGAGGAGAAGCAGGAT 60

Qy    167 ACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAA 226
          || |||||||||||||||||| ||||||| |||||| || ||||||||||||||||||
Db     61 ACACAGGAGAAAGAAGGTATTCTTCCTGAGAAAGCTGAGGAAGCAAAGCTAAAGGCCAAA 120

Qy    227 TACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAA 286
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TATCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAA 180

Qy    287 GGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAG 346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAG 240

Qy    347 CAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCA 406
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    241 CAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGACCACATCCCCACCCCA 300

Qy    407 CAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 350


RESULT  18
BTA005986
LOCUS       BTA005986               366 bp    mRNA    linear   MAM 07-JAN-1999
DEFINITION  Bos taurus alpha-endosulfine (complete coding sequence).
ACCESSION   AJ005986
VERSION     AJ005986.1  GI:4127577
KEYWORDS    alpha-endosulfine.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Bataille D., U 376 'Endocrinologie des
            Peptides et regulation genique', INSERM, CHU Arnaud-de-Villeneuve,
            34295, FRANCE
REFERENCE   2  (bases 1 to 366)
  AUTHORS   Virsolvy-Vergine,A., Salazar,G., Sillard,R., Denoroy,L., Mutt,V.
            and Bataille,D.
  TITLE     Endosulfine, endogenous ligand for the sulphonylurea receptor:
            isolation from porcine brain and partial structural determination
            of the alpha form
  JOURNAL   Diabetologia 39 (2), 135-141 (1996)
  MEDLINE   96218590
REFERENCE   3  (bases 1 to 366)
  AUTHORS   Peyrollier,K., Heron,L., Virsolvy-Vergine,A., Le Cam,A. and
            Bataille,D.
  TITLE     Alpha endosulfine is a novel molecule, structurally related to a
            family of phosphoproteins
  JOURNAL   Biochem. Biophys. Res. Commun. 223 (3), 583-586 (1996)
  MEDLINE   96268722
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /tissue_type="brain"
     CDS             1. .366
                     /function="endogenous regulator of K-ATP channels"
                     /codon_start=1
                     /evidence=experimental
                     /product="alpha-endosulfine"
                     /protein_id="CAA06800.1"
                     /db_xref="GI:4127578"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPEKAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAGGQVE"
BASE COUNT      129 a     94 c     96 g     47 t
ORIGIN


  Query Match           30.1%;  Score 332.4;  DB 4;  Length 366;
  Best Local Similarity   96.9%;  Pred. No. 1.2e-86;
  Matches  339;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;
```

```
Qy    107 ATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGAC 166
          ||||||||||||||||||||| ||||||||||||| ||||||||||||||||||||
Db      1 ATGTCCCAGAAACAAGAAGAAGAAAACCCTGCGGAGGAAACCGGCGAGGAGAAGCAGGAT 60

Qy    167 ACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAA 226
          || |||||||||||||||||| ||||||| |||||| || ||||||||||||||||||
Db     61 ACACAGGAGAAAGAAGGTATTCTTCCTGAGAAAGCTGAGGAAGCAAAGCTAAAGGCCAAA 120

Qy    227 TACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAA 286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAA 180

Qy    287 GGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAG 346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAG 240

Qy    347 CAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCA 406
          |||||||||||||||||||||||||||||||||||| ||||| ||||||||||||||||
Db    241 CAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACCGGTGACCACATCCCCACCCCA 300

Qy    407 CAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
          ||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    301 CAGGACCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 350


RESULT  19
MMA005985
LOCUS       MMA005985               366 bp    mRNA    linear   ROD 07-JAN-1999
DEFINITION  Mus musculus alpha-endosulfine (complete coding sequence).
ACCESSION   AJ005985
VERSION     AJ005985.1  GI:4138191
KEYWORDS    alpha-endosulfine.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Bataille D., U 376 'Endocrinologie des
            Peptides et regulation genique', INSERM, CHU Arnaud-de-Villeneuve,
            34295, FRANCE
REFERENCE   2  (bases 1 to 366)
  AUTHORS   Virsolvy-Vergine,A., Salazar,G., Sillard,R., Denoroy,L., Mutt,V.
            and Bataille,D.
  TITLE     Endosulfine, endogenous ligand for the sulphonylurea receptor:
            isolation from porcine brain and partial structural determination
            of the alpha form
  JOURNAL   Diabetologia 39 (2), 135-141 (1996)
  MEDLINE   96218590
REFERENCE   3  (bases 1 to 366)
  AUTHORS   Peyrollier,K., Heron,L., Virsolvy-Vergine,A., Le Cam,A. and
            Bataille,D.
  TITLE     Alpha endosulfine is a novel molecule, structurally related to a
            family of phosphoproteins
  JOURNAL   Biochem. Biophys. Res. Commun. 223 (3), 583-586 (1996)
  MEDLINE   96268722
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /tissue_type="brain"
     CDS             1. .366
                     /function="endogenous regulator of K-ATP channels"
                     /codon_start=1
                     /evidence=experimental
                     /product="alpha-endosulfine"
                     /protein_id="CAA06799.1"
                     /db_xref="GI:4138192"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPEKAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGADKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAGGQVE"
BASE COUNT     127 a     94 c    100 g     45 t
```

```
ORIGIN


  Query Match           30.1%;  Score 332.4;  DB 10;  Length 366;
  Best Local Similarity  96.9%;  Pred. No. 1.2e-86;
  Matches  339;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy    107 ATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGAC 166
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      1 ATGTCCCAGAAACAAGAAGAAGAAAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGAT 60

Qy    167 ACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAA 226
          || |||||||||||||| ||||| ||||||| |||||| ||||||||||||||||||||
Db     61 ACACAGGAGAAAGAAGGGATTCTCCCTGAGAAAGCTGAGGAGGCAAAGCTAAAGGCCAAA 120

Qy    227 TACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAA 286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAA 180

Qy    287 GGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAG 346
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    181 GGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAACAAG 240

Qy    347 CAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCA 406
          |||||||||||||||||| |||||||||||||||||| ||||| ||||||||||||||||
Db    241 CAGCTGCCAAGTGCAGGAGCAGACAAGAACCTGGTGACCGGTGACCACATCCCCACCCCA 300

Qy    407 CAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 350


RESULT  20
RNA005984
LOCUS       RNA005984               366 bp    mRNA    linear   ROD 07-JAN-1999
DEFINITION  Rattus norvegicus mRNA alpha-endosulfine (complete coding
            sequence).
ACCESSION   AJ005984
VERSION     AJ005984.1  GI:4138425
KEYWORDS    Rattus norvegicus.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Bataille D., U 376 'Endocrinologie des
            Peptides et regulation genique', INSERM, CHU Arnaud-de-Villeneuve,
            34295, FRANCE
REFERENCE   2  (bases 1 to 366)
  AUTHORS   Virsolvy-Vergine,A., Salazar,G., Sillard,R., Denoroy,L., Mutt,V.
            and Bataille,D.
  TITLE     Endosulfine, endogenous ligand for the sulphonylurea receptor:
            isolation from porcine brain and partial structural determination
            of the alpha form
  JOURNAL   Diabetologia 39 (2), 135-141 (1996)
  MEDLINE   96218590
REFERENCE   3  (bases 1 to 366)
  AUTHORS   Peyrollier,K., Heron,L., Virsolvy-Vergine,A., Le Cam,A. and
            Bataille,D.
  TITLE     Alpha endosulfine is a novel molecule, structurally related to a
            family of phosphoproteins
  JOURNAL   Biochem. Biophys. Res. Commun. 223 (3), 583-586 (1996)
  MEDLINE   96268722
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /tissue_type="brain"
     CDS             1. .366
                     /function="endogenous regulator of K-ATP channels"
```

```
                     /note="Isolated as an endogenous ligand for the
                     sulphonylurea receptor"
                     /codon_start=1
                     /evidence=experimental
                     /product="alpha-endosulfine"
                     /protein_id="CAA06798.1"
                     /db_xref="GI:4138426"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPEKAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGADKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAGGQVE"
BASE COUNT      129 a     90 c     97 g     50 t
ORIGIN


  Query Match             29.8%;  Score 329.2;  DB 10;  Length 366;
  Best Local Similarity   96.3%;  Pred. No. 1e-85;
  Matches  337;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy     107 ATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGAC 166
           |||||||||||||||||||||| |||||||||||||| ||||||||||||||||||||
Db       1 ATGTCCCAGAAACAAGAAGAAGAAAACCCTGCGGAGGAAACCGGCGAGGAGAAGCAGGAT 60

Qy     167 ACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAA 226
           || |||||||||||||||||| ||||||| |||||| || ||||||||||||||||||
Db      61 ACACAGGAGAAAGAAGGTATTCTTCCTGAGAAAGCTGAGGAAGCAAAGCTAAAGGCCAAA 120

Qy     227 TACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAA 286
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TATCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAA 180

Qy     287 GGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAG 346
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     181 GGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAACAAG 240

Qy     347 CAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCA 406
           ||||| || ||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     241 CAGCTACCGAGTGCAGGAGCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCA 300

Qy     407 CAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 350



RESULT  21
AC092479/c
LOCUS       AC092479               142023 bp    DNA     linear   HTG 05-FEB-2002
DEFINITION  Mus musculus clone rp23-231l15 strain C57BL/6J, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC092479
VERSION     AC092479.12  GI:18497125
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 142023)
  AUTHORS   Jia,H., Zhang,P., Lin,S., Wu,H. and Roe,B.A.
  TITLE     Mus musculus BAC Clone rp23-231l15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 142023)
  AUTHORS   Jia,H., Zhang,P., Lin,S., Wu,H. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2001) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Feb 5, 2002 this sequence version replaced gi:18390246.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    32949: contig of 32949 bp in length
            *   32950    33049: gap of unknown length
            *   33050    78918: contig of 45869 bp in length
            *   78919    79018: gap of unknown length
            *   79019   142023: contig of 63005 bp in length.
FEATURES             Location/Qualifiers
     source          1. .142023
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="rp23-231l15"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    37523 a  34071 c  31510 g  38713 t    206 others
ORIGIN


  Query Match             29.1%;  Score 321;  DB 2;  Length 142023;
  Best Local Similarity   72.5%;  Pred. No. 1.1e-82;
  Matches  582;  Conservative    0;  Mismatches  175;  Indels   46;  Gaps   11;

Qy     289 GCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCA 348
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db 137894 GCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAACAAGCA 137835

Qy     349 GCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACA 408
           |||||||||||||||| |||||||||||||||||| ||||| ||||||||||||||||
Db 137834 GCTGCCAAGTGCAGGAGCAGACAAGAACCTGGTGACCGGTGACCACATCCCCACCCCACA 137775

Qy     409 GGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCC 468
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db 137774 GGACCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGGG-- 137715

Qy     469 CCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATG 528
           |||||||| |||||||||||| ||||||||| ||||||||| |||||||||    ||||
Db 137716 CCCCTCTCTTCCCCTTCCTCACCCACTGGACTGTTATATATTGTAGGCAGGG----AATG 137661

Qy     529 GGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGC 588
           ||||||||||||| ||          ||| ||||   | |   | ||||||||||||||
Db 137660 GGCACCTAGTCAGAGCT-------CAGCTTGCCATCCAGGGGAGACATTCTGCTGGGGGC 137608

Qy     589 TGCTGAGAAGGTAATGTAGGTTGAAAAGGGCTCTAAGTTTATTTATTTCGTTAGATTGA 648
           |               ||||||||| | ||||||| |||| |||| ||| | || |||||
Db 137607 T---------------TAGGTTGAAGATGGGCTCTGAGTTCATTTCTTTGGGTAAATTGA 137563

Qy     649 CACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGG 708
           |||| |||| |||    ||| ||||    | || |  | |||||||| ||||||| ||
Db 137562 AACTTGTACACCCTCATGATAGCCCAGTAGAGACCCAAGAGTAGGAAAGACTAGGAGTGC 137503

Qy     709 ATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGG 768
           | | |||||||   |||||| ||||     ||   |  |||||  | ||||||| ||||
Db 137502 AGAGTGCTGCA---GCTTTTCCTGTGGTGAGAATGTGGAAGATGCAAGTTCTCCTGTTGG 137446

Qy     769 GAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACAT 828
           |||  | ||||  | |  |  ||   |         |||  ||| || ||||||| |   |
Db 137445 AGGAGCAAGTCCTTAGACGGGCTATCGT--------GGAGCAGTGTGACAGGTTCAGTGT 137394

Qy     829 CCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCT 888
           |||||||| |  | || | | || |||||||||| |||| |||  ||| | ||||||||
Db 137393 CCCAAGGCTGAATTGACAT-GCTGAGAAAGGTAGAGTGAAGGAG--ATGCTCTCTGTGCT 137337

Qy     889 TCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGA 948
           ||||||||| ||||||||||||  || |    ||    |||| |   |  ||||||
Db 137336 TCTAAGTGTGCTGTCCCTTAGGC--CTCTGTGCTCTGTGTTCCCATCGTGGGTGTTTAGG 137279

Qy     949 GAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCC 1008
           | | | | ||||   | |     ||||||| ||||| ||||| |||| | || || ||
Db 137278 GGACCTTCAAAAGAGAGGAGATTTCTGCTTGCTTAAAATTACAGGGCCCACCTAACC-CC 137220
```

```
Qy   1009 ATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACA 1068
          | ||  ||  |||  ||||  ||| |||  ||||||  | | ||||| || |   ||
Db 137219 TCTGCTGTGTTCCTCCAACAG-CTGCCCTGCTTCTCAAGGAGTGTAAAGTTTAACAGACG 137161

Qy   1069 TTATTCACAAAAAAAAAAAAAAA 1091
          |||   ||  |  ||||| ||||
Db 137160 TTAGCTACTTATGAAAAAGAAAA 137138
```

RESULT 22
HSJ822J19/c
LOCUS HSJ822J19 89872 bp DNA linear PRI 19-MAR-2001
DEFINITION Human DNA sequence from clone RP5-822J19 on chromosome 20. Contains an alpha-endosulfine pseudogene, STSs and GSSs, complete sequence.
ACCESSION AL050324
VERSION AL050324.5 GI:5002694
KEYWORDS HTG; endosulfine.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 89872)
AUTHORS Matthews,L.
TITLE Direct Submission
JOURNAL Submitted (14-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Jun 7, 1999 this sequence version replaced gi:4972283.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr20
RP5-822J19 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP5-822J19.

FEATURES Location/Qualifiers

```
     source          1. .89872
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP5-822J19"
                     /clone_lib="RPCI-5"
     repeat_region   1. .112
                     /note="MER5B repeat: matches 67. .178 of consensus"
     misc_feature    113. .601
                     /note="match: STS: Em:HS822J19S"
     repeat_region   833. .865
                     /note="MER5A repeat: matches 14. .46 of consensus"
     repeat_region   1006. .1313
                     /note="AluSg repeat: matches 1. .308 of consensus"
     repeat_region   1349. .1679
                     /note="LTR16C repeat: matches 53. .377 of consensus"
     repeat_region   1979. .2170
                     /note="MER5A repeat: matches 1. .186 of consensus"
     repeat_region   2526. .2859
                     /note="L1M4 repeat: matches 4254. .4604 of consensus"
     repeat_region   2862. .3211
                     /note="THE1A repeat: matches 1. .354 of consensus"
     repeat_region   3212. .4798
                     /note="THE1A-internal repeat: matches 1. .1580 of consensus"
     repeat_region   4799. .5152
                     /note="THE1A repeat: matches 1. .354 of consensus"
     repeat_region   5194. .5680
                     /note="L1MC1 repeat: matches 5813. .6317 of consensus"
     repeat_region   6026. .6232
                     /note="MIR repeat: matches 35. .247 of consensus"
     repeat_region   6419. .6715
                     /note="AluJb repeat: matches 1. .309 of consensus"
     repeat_region   6727. .6850
                     /note="31 copies 4 mer aaga 76% conserved"
     repeat_region   6737. .6850
                     /note="57 copies 2 mer ga 75% conserved"
     repeat_region   7405. .7434
                     /note="15 copies 2 mer aa 100% conserved"
     repeat_region   7859. .8091
                     /note="MIR repeat: matches 16. .249 of consensus"
     misc_feature    complement(8145. .8666)
                     /note="match: GSS: Em:AQ603396"
     misc_feature    complement(8339. .8673)
                     /note="match: STS: Em:HSPF05H8"
     misc_feature    complement(9561. .10143)
                     /note="match: GSS: Em:AQ582705"
     repeat_region   10445. .10777
                     /note="L1 repeat: matches 4859. .5202 of consensus"
     repeat_region   10928. .11110
                     /note="L1MB3 repeat: matches 5981. .6181 of consensus"
     repeat_region   11722. .11864
                     /note="MIR repeat: matches 112. .260 of consensus"
     repeat_region   13615. .13909
                     /note="AluSx repeat: matches 1. .294 of consensus"
     repeat_region   14917. .15218
                     /note="AluSc repeat: matches 1. .289 of consensus"
     repeat_region   15341. .15443
                     /note="MIR repeat: matches 72. .197 of consensus"
     repeat_region   17565. .17859
                     /note="AluY repeat: matches 1. .295 of consensus"
     repeat_region   17994. .18333
                     /note="LTR16A repeat: matches 96. .447 of consensus"
     repeat_region   19036. .19340
                     /note="AluSx repeat: matches 1. .300 of consensus"
     repeat_region   20208. .20306
                     /note="MER5A repeat: matches 86. .187 of consensus"
     repeat_region   21264. .21315
                     /note="13 copies 4 mer gtgt 86% conserved"
     repeat_region   21265. .21314
                     /note="25 copies 2 mer tg 88% conserved"
     repeat_region   21844. .22314
                     /note="L2 repeat: matches 1997. .2502 of consensus"
     repeat_region   22539. .22839
                     /note="AluSp repeat: matches 1. .302 of consensus"
     repeat_region   22854. .22984
                     /note="FLAM_C repeat: matches 3. .133 of consensus"
     repeat_region   22985. .23018
                     /note="BC200 repeat: matches 1. .161 of consensus"
     misc_feature    complement(23059. .23343)
                     /note="match: STS: Em:G07642"
     repeat_region   23344. .23651
                     /note="AluSx repeat: matches 1. .308 of consensus"
     misc_feature    complement(23652. .23753)
                     /note="match: GSS: Em:AQ807387"
     misc_feature    complement(23652. .23754)
                     /note="match: GSS: Em:AQ544217"
     misc_feature    complement(23652. .23752)
                     /note="match: GSS: Em:AQ544233"
```

```
     misc_feature    23754. .24328
                     /note="match: GSS: Em:AQ346979"
     misc_feature    23900. .24386
                     /note="match: GSS: Em:AQ711642"
     repeat_region   24606. .24731
                     /note="63 copies 2 mer tt 59% conserved"
     misc_feature    complement(24848. .25354)
                     /note="match: GSS: Em:AQ561985"
     misc_feature    complement(24849. .25354)
                     /note="match: GSS: Em:AQ561966"
     misc_feature    25355. .25824
                     /note="match: GSS: Em:AQ345431"
     misc_feature    25358. .26074
                     /note="match: GSS: Em:AQ744401"
     misc_feature    25369. .25875
                     /note="match: GSS: Em:AQ693363"
     repeat_region   26075. .26496
                     /note="MLT1C repeat: matches 1. .438 of consensus"
     repeat_region   26808. .26833
                     /note="13 copies 2 mer tt 96% conserved"
     repeat_region   26953. .27167
                     /note="MIR repeat: matches 36. .250 of consensus"
     misc_feature    complement(29003. .29207)
                     /note="match: GSS: Em:AQ101496"
     repeat_region   29992. .30051
                     /note="MIR repeat: matches 195. .254 of consensus"
     misc_feature    31069. .31571
                     /note="match: GSS: Em:AQ371017"
     misc_feature    31071. .31481
                     /note="match: GSS: Em:AQ424527"
     misc_feature    31073. .31596
                     /note="match: GSS: Em:AQ568767"
     repeat_region   31914. .32221
                     /note="AluSx repeat: matches 1. .307 of consensus"
     repeat_region   33461. .33764
                     /note="L2 repeat: matches 2379. .2750 of consensus"
     repeat_region   33912. .33935
                     /note="12 copies 2 mer tt 95% conserved"
     repeat_region   34564. .35249
                     /note="LTR8 repeat: matches 1. .691 of consensus"
     gene            complement(35297. .35637)
                     /gene="dJ822J19.1"
     CDS             complement(35297. .35637)
                     /gene="dJ822J19.1"
                     /note="dJ822J19.1 (alpha-endosulfine pseudogene)
                     match: proteins: Tr:Q9Z0G1 Tr:O97976 Sw:P56212 Sw:Q28055
                     Tr:P79322 Tr:O43768"
                     /codon_start=1
                     /pseudo
                     /evidence=not_experimental
     repeat_region   35829. .35885
                     /note="L2 repeat: matches 2036. .2093 of consensus"
     misc_feature    36012. .36251
                     /note="match: GSS: Em:AQ243178"
     repeat_region   36252. .36402
                     /note="MIR repeat: matches 99. .262 of consensus"
     repeat_region   36590. .36617
                     /note="7 copies 4 mer tgtt 96% conserved"
     repeat_region   38149. .38220
                     /note="MER44 repeat: matches 657. .728 of consensus"
     repeat_region   38229. .38490
                     /note="MER44 repeat: matches 396. .653 of consensus"
     repeat_region   38548. .38816
                     /note="MER44A repeat: matches 1. .266 of consensus"
     repeat_region   39491. .39566
                     /note="19 copies 4 mer tatc 90% conserved"

  Query Match             25.3%;  Score 278.6;  DB 9;  Length 89872;
  Best Local Similarity   81.8%;  Pred. No. 3.1e-70;
  Matches  372;  Conservative    0;  Mismatches   74;  Indels    9;  Gaps    4;

Qy       2 TGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCC 61
           |||||||  || ||||||||| ||||||| |||||||  || |||  ||| | || |   |

Db   35758 TGACAGGGACCAAAGCAGCATCGCAGGATGTCCCGTTTCCCCTCCCCACCTCTCAGTT-- 35699

Qy      62 GGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAA 121
           ||||| ||||| || || | || |||| |  ||||| ||||| | ||||||||||||
Db   35700 GGTTGACTTCCTGGACCTCCTATACTCTATTGTCCCAGTCCCACAATGTCCCAGAAAC-- 35641

Qy     122 GAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAA 181
             | | |||||| | | ||||||||| ||| || |||||||||| || || ||||||||||
Db   35642 -TAAACGAGAACACCGTGGAGGAGACAGGCAAGAAGAAGCAGGGCATGCGGGAGAAAGAA 35584

Qy     182 GGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGA 241
           ||| ||| ||||||| ||  ||  ||| |||||||||||||||||| ||||||||||||
Db   35583 TGTACTCTCCCTGAGACAGTGGAGAAGGAAAAGCTAAAGGCCAAATATCCAAGCCTAGGA 35524

Qy     242 CAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTT 301
           |||| |||||| ||| ||  |||||| |||||| ||||||  |||||| ||||||||||
Db   35523 CAAAGGCCTGGGGGC-CCAGCTTCCTTATGAAGGGACTCCGAAAAGGGAAAAAGTACTTT 35465

Qy     302 GACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCA 361
           | |||||||||||||||||||||||||||||||    | ||||||||||||||||||||||
Db   35464 GCCTCAGGAGACTACAACATGGCCAAAGCCAAATAGGAGAATAAGCAGCTGCCAAGTGCA 35405

Qy     362 GGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAG 421
           ||||||||||||||| |||||||||| |||||||||||| ||||||||||||||
Db   35404 CAACCAGACAAGAACCTGATGACTGGTGACCACATCCCCACCACACAGGATCTGCCC--- 35346

Qy     422 AGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
           ||||| ||||| ||| ||||||||||||  |||||
Db   35347 AGAAAATCCTCACTCATCACCAGCAAGCCAGCGGG 35313


RESULT  23
AC007124/c
LOCUS       AC007124               89263 bp    DNA     linear   PRI 20-MAR-1999
DEFINITION  Human genomic clone P1-11874, complete sequence.
ACCESSION   AC007124
VERSION     AC007124.1  GI:4454586
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 89263)
  AUTHORS   Nascimento,L., Spiegel,L.A., de la Bastide,M., Habermann,K.,
            See,L.H., Rodriguez,M., O'Shaughnessy,A., Matero,A., Swaby,I.,
            Huang,E.N., Schutz,K., Preston,R.R., Shah,R., Vil,M.D., Shekher,M.,
            Parnell,L.D., Dedhia,N.N. and McCombie,W.R.
  TITLE     Homo sapiens genomic clone P1-11874
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 89263)
  AUTHORS   Nascimento,L., Spiegel,L.A., de la Bastide,M., Habermann,K.,
            See,L.H., Rodriguez,M., O'Shaughnessy,A., Matero,A., Swaby,I.,
            Huang,E.N., Schutz,K., Preston,R.R., Shah,R., Vil,M.D., Shekher,M.,
            Parnell,L.D., Dedhia,N.N. and McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-1999) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724
COMMENT     Position 1 of P1-11874 is at position 16011 of cloning vector
            pSacBII and position 89263 is at position 1 of pSacBII. P1-11874 is
            40.6% G+C.
FEATURES             Location/Qualifiers
     source          1. .89263
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT    25953 a  18074 c  18201 g  27035 t
ORIGIN


  Query Match             25.1%;  Score 277;  DB 9;  Length 89263;
  Best Local Similarity   81.5%;  Pred. No. 9.2e-70;
  Matches  371;  Conservative    0;  Mismatches   75;  Indels    9;  Gaps    4;
```

```
Qy       2 TGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCC 61
           ||||||| || |||||||| |||||| ||||||  || |||  ||| | || |   |
Db  44421 TGACAGGGACCAAAGCAGCATCGCAGGATGTCCCGTTTCCCCTCCCCACCTCTCAGTT-- 44362

Qy      62 GGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAA 121
           ||||| ||||| || || | || |||| |  ||||| ||||| | ||||||||||||
Db  44363 GGTTGACTTCCTGGACCTCCTATACTCTATTGTCCCAGTCCCACAATGTCCCAGAAAC-- 44304

Qy     122 GAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAA 181
             | | |||||| | | ||||||||| ||| || ||||||||| || || ||||||||||
Db  44305 -TAAACGAGAACACCGTGGAGGAGACAGGCAAGAAGAAGCAGGGCATGCGGGAGAAAGAA 44247

Qy     182 GGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGA 241
           ||| ||| ||||||| ||  ||  ||| |||||||||||||||| ||||||||||
Db  44246 TGTACTCTCCCTGAGACAGTGGAGAAGGAAAAGCTAAAGGCCAAATATCCAAGCCTAGGA 44187

Qy     242 CAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTT 301
           |||| |||||| ||| ||  |||||| |||||| ||||||  |||||| ||||||||||
Db  44186 CAAAGGCCTGGGGGC-CCAGCTTCCTTATGAAGGGACTCCGAAAAGGGAAAAAGTACTTT 44128

Qy     302 GACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCA 361
           | ||||||||||||||||||||||||||||   | ||||||||||||||||||||||
Db  44127 GCCTCAGGAGACTACAACATGGCCAAAGCCAAATAGGAGAATAAGCAGCTGCCAAGTGCA 44068

Qy     362 GGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAG 421
           ||||||||||||||| |||||||||| |||||||||||| |||||||||||
Db  44067 CAACCAGACAAGAACCTGATGACTGGTGACCACATCCCCACCACACAGGATCTGCCC--- 44009

Qy     422 AGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
           ||||| ||||| ||| |||||||||||  | |||
Db  44010 AGAAAATCCTCACTCATCACCAGCAAGCCAGTGGG 43976
```

```
RESULT  24
AC007449
LOCUS       AC007449              176616 bp    DNA     linear   HTG 13-APR-2000
DEFINITION  Homo sapiens clone RP11-104A1, WORKING DRAFT SEQUENCE, 13 unordered
            pieces.
ACCESSION   AC007449
VERSION     AC007449.2  GI:7547180
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176616)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-104A1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176616)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 13, 2000 this sequence version replaced gi:4726138.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L773
                Center clone name: 104_A_1
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 91% of reads
                Sequencing vector: Plasmid; n/a; %-0.f%% of reads
                8.87005649717514Chemistry: Dye-primer-amersham; 97% of reads
                Chemistry: Dye-terminator Big Dye; 3% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 170592 bases at least Q40
                Consensus quality: 173409 bases at least Q30
                Consensus quality: 174613 bases at least Q20
                Insert size: 180000; agarose-fp
                Insert size: 175416; sum-of-contigs
                Quality coverage.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1       51: contig of 51 bp in length
            *       52 151: gap of      100 bp
            *      152     1681: contig of 1530 bp in length
            *     1682 1781: gap of      100 bp
            *     1782     5071: contig of 3290 bp in length
            *     5072 5171: gap of      100 bp
            *     5172     9359: contig of 4188 bp in length
            *     9360 9459: gap of      100 bp
            *     9460    13792: contig of 4333 bp in length
            *    13793 13892: gap of      100 bp
            *    13893    23757: contig of 9865 bp in length
            *    23758 23857: gap of      100 bp
            *    23858    38876: contig of 15019 bp in length
            *    38877 38976: gap of      100 bp
            *    38977    52701: contig of 13725 bp in length
            *    52702 52801: gap of      100 bp
            *    52802    73312: contig of 20511 bp in length
            *    73313 73412: gap of      100 bp
            *    73413    92674: contig of 19262 bp in length
            *    92675 92774: gap of      100 bp
            *    92775   114928: contig of 22154 bp in length
            *   114929 115028: gap of      100 bp
            *   115029   143307: contig of 28279 bp in length
            *   143308 143407: gap of      100 bp
            *   143408   176616: contig of 33209 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176616
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-104A1"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .51
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    152. .1681
                     /note="assembly_fragment"
     misc_feature    1782. .5071
                     /note="assembly_fragment"
     misc_feature    5172. .9359
                     /note="assembly_fragment"
     misc_feature    9460. .13792
                     /note="assembly_fragment"
     misc_feature    13893. .23757
                     /note="assembly_fragment"
     misc_feature    23858. .38876
```

```
                     /note="assembly_fragment"
     misc_feature    38977. .52701
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    52802. .73312
                     /note="assembly_fragment"
     misc_feature    73413. .92674
                     /note="assembly_fragment"
     misc_feature    92775. .114928
                     /note="assembly_fragment"
     misc_feature    115029. .143307
                     /note="assembly_fragment"
     misc_feature    143408. .176616
                     /note="assembly_fragment"
BASE COUNT    52499 a  35498 c  34632 g  52787 t   1200 others
ORIGIN


  Query Match           25.1%;  Score 276.6;  DB 2;  Length 176616;
  Best Local Similarity  81.5%;  Pred. No. 1.4e-69;
  Matches  371;  Conservative    0;  Mismatches   74;  Indels   10;  Gaps    4;

Qy       2 TGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCC 61
           ||||||| || |||||||| |||||| |||||  || |||  ||| | || |   |
Db   59902 TGACAGGGACCAAAGCAGCATCGCAGGATGTCCCGTTTCCCCTCCCCACCTCTCAGTT-- 59959

Qy      62 GGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAA 121
           ||||| ||||| || || | || |||| |  ||||| ||||| | ||||||||||||
Db   59960 GGTTGACTTCCTGGACCTCCTATACTCTATTGTCCCAGTCCCACAATGTCCCAGAAAC-- 60017

Qy     122 GAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAA 181
             | | |||||| | | ||||||||| ||| || |||||||||  || || ||||||||||
Db   60018 -TAAACGAGAACACCGTGGAGGAGACAGGCAAGAAGAAGCAGGGCATGCGGGAGAAAGAA 60076

Qy     182 GGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGA 241
           ||| ||| |||||||  ||  ||  ||| |||||||||||||||||  ||||||||||||
Db   60077 TGTACTCTCCCTGAGACAGTGGAGAAGGAAAAGCTAAAGGCCAAATATCCAAGCCTAGGA 60136

Qy     242 CAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTT 301
           |||| |||||| ||  ||  |||||| |||||| ||||||  |||||| ||||||||||
Db   60137 CAAAGGCCTGGGGG--CCAGCTTCCTTATGAAGGGACTCCGAAAAGGGAAAAAGTACTTT 60194

Qy     302 GACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCA 361
           | ||||||||||||||||||||||||||||||    | ||||||||||||||||||||||
Db   60195 GCCTCAGGAGACTACAACATGGCCAAAGCCAAATAGGAGAATAAGCAGCTGCCAAGTGCA 60254

Qy     362 GGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAG 421
             |||||||||||||| |||||||||| ||||||||||| ||||||||||||||
Db   60255 CAACCAGACAAGAACCTGATGACTGGTGACCACATCCCCACCACACAGGATCTGCCC--- 60311

Qy     422 AGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
           ||||| ||||| ||| ||||||||||||  |||||
Db   60312 AGAAAATCCTCACTCATCACCAGCAAGCCAGCGGG 60346


RESULT  25
AC103421
LOCUS       AC103421              105126 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-148B7, *** SEQUENCING IN PROGRESS
            ***; 47 unordered pieces.
ACCESSION   AC103421
VERSION     AC103421.2  GI:17973226
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 105126)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 105126)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 20, 2001 this sequence version replaced gi:17064207.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GHMB
                Center clone name: CH230-148B7
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 87267 bases at least Q40
                Consensus quality: 93895 bases at least Q30
                Consensus quality: 100375 bases at least Q20
                Estimated insert size: 85832; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 47 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

```
*       1     4645: contig of 4645 bp in length
*    4646     4745: gap of unknown length
*    4746     8474: contig of 3729 bp in length
*    8475     8574: gap of unknown length
*    8575    12295: contig of 3721 bp in length
*   12296    12395: gap of unknown length
*   12396    17604: contig of 5209 bp in length
*   17605    17704: gap of unknown length
*   17705    21282: contig of 3578 bp in length
*   21283    21382: gap of unknown length
*   21383    24900: contig of 3518 bp in length
*   24901    25000: gap of unknown length
*   25001    27522: contig of 2522 bp in length
*   27523    27622: gap of unknown length
*   27623    30513: contig of 2891 bp in length
*   30514    30613: gap of unknown length
*   30614    34148: contig of 3535 bp in length
*   34149    34248: gap of unknown length
*   34249    37150: contig of 2902 bp in length
*   37151    37250: gap of unknown length
*   37251    40253: contig of 3003 bp in length
*   40254    40353: gap of unknown length
*   40354    42860: contig of 2507 bp in length
*   42861    42960: gap of unknown length
*   42961    45644: contig of 2684 bp in length
*   45645    45744: gap of unknown length
*   45745    47445: contig of 1701 bp in length
*   47446    47545: gap of unknown length
*   47546    49806: contig of 2261 bp in length
*   49807    49906: gap of unknown length
*   49907    53212: contig of 3306 bp in length
*   53213    53312: gap of unknown length
*   53313    56240: contig of 2928 bp in length
*   56241    56340: gap of unknown length
*   56341    59832: contig of 3492 bp in length
*   59833    59932: gap of unknown length
*   59933    62010: contig of 2078 bp in length
*   62011    62110: gap of unknown length
*   62111    63482: contig of 1372 bp in length
*   63483    63582: gap of unknown length
*   63583    65587: contig of 2005 bp in length
*   65588    65687: gap of unknown length
*   65688    67278: contig of 1591 bp in length
*   67279    67378: gap of unknown length
*   67379    69837: contig of 2459 bp in length
*   69838    69937: gap of unknown length
*   69938    71641: contig of 1704 bp in length
*   71642    71741: gap of unknown length
*   71742    72772: contig of 1031 bp in length
*   72773    72872: gap of unknown length
*   72873    74038: contig of 1166 bp in length
*   74039    74138: gap of unknown length
*   74139    75347: contig of 1209 bp in length
*   75348    75447: gap of unknown length
*   75448    76582: contig of 1135 bp in length
*   76583    76682: gap of unknown length
*   76683    78565: contig of 1883 bp in length
*   78566    78665: gap of unknown length
*   78666    80058: contig of 1393 bp in length
*   80059    80158: gap of unknown length
*   80159    81800: contig of 1642 bp in length
*   81801    81900: gap of unknown length
*   81901    83526: contig of 1626 bp in length
*   83527    83626: gap of unknown length
*   83627    85206: contig of 1580 bp in length
*   85207    85306: gap of unknown length
*   85307    86373: contig of 1067 bp in length
*   86374    86473: gap of unknown length
*   86474    88280: contig of 1807 bp in length
*   88281    88380: gap of unknown length
*   88381    90567: contig of 2187 bp in length
*   90568    90667: gap of unknown length
*   90668    91832: contig of 1165 bp in length
*   91833    91932: gap of unknown length
*   91933    92951: contig of 1019 bp in length
*   92952    93051: gap of unknown length
*   93052    94177: contig of 1126 bp in length
*   94178    94277: gap of unknown length
*   94278    95584: contig of 1307 bp in length
*   95585    95684: gap of unknown length
*   95685    96966: contig of 1282 bp in length
*   96967    97066: gap of unknown length
*   97067    98319: contig of 1253 bp in length
*   98320    98419: gap of unknown length
*   98420    99500: contig of 1081 bp in length
*   99501    99600: gap of unknown length
*   99601   100809: contig of 1209 bp in length
*  100810   100909: gap of unknown length
*  100910   102534: contig of 1625 bp in length
*  102535   102634: gap of unknown length
*  102635   103650: contig of 1016 bp in length
*  103651   103750: gap of unknown length
*  103751   105126: contig of 1376 bp in length.
FEATURES             Location/Qualifiers
     source          1. .105126
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-148B7"
BASE COUNT    31209 a  19023 c  19380 g  30766 t   4748 others
ORIGIN


  Query Match             20.7%;  Score 228.2;  DB 2;  Length 105126;
  Best Local Similarity   69.5%;  Pred. No. 2.3e-55;
  Matches  380;  Conservative    0;  Mismatches  138;  Indexes   29;  Gaps    4;

Qy     173 GAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCA 232
           ||| ||| |||  | ||||| || || | ||  | ||||| ||| |||
Db   97782 GAGGGAGAGAAGGATCTAGGAGTGAGAGAGGATGAAGTTAAAAGGAGGGCCATATATCCA 97841

Qy     233 AGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAA 292
           |||| ||||||||||||||||||  | |||||||| |||||||  || |||||||
Db   97842 CATCTAGAACAAAAGCCTGGAGGCTCCATCATCCTCATGGAGAGACTCTTGATAGGGCAA 97901

Qy     293 AAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTG 352
           || |||||||| |||||||||| | || |  ||||||||||||||||  |||||| ||
Db   97902 AATTACTTTGAATCAGGAGACTATACCAAGTTCAAAGCCAAGATGAAGATCAAGCAGATG 97961

Qy     353 CCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGAT 412
           |  |||||||  |||||| |||| || ||||  |  | | |||||| | | ||||||||
Db   97962 CTGAGTGCAGGGGCAGACATGAACTTGATGACCAGCAACCTCATCCCTATCACACAGGAT 98021

Qy     413 CTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCC 472
           |||||||||||| ||| ||| ||  ||| |||||||||| ||| ||||||  |   |||
Db   98022 CTGCCCCAGAGAGAGTTCTCACTTATCATCAGCAAGCTTGTGGGATAACCT--GGATCCC 98079

Qy     473 TCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCA 532
           |||| ||||| |||||| |||||||| ||||||| ||  ||||||||     |||||||
Db   98080 TCTCTTCCCCATCCTCACCCACTGGACTTTTATATGTTGAAGGCAGGG-----ATGGGCA 98134

Qy     533 CCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCT 592
           |||||||||  ||       |  || ||||  | |   |  ||| ||||| | |||
Db   98135 CCTAGTCAGTGCT-------TAACTTGCCATCCAGGGAAGACATTGTGCTGAGTGCTC-- 98185

Qy     593 GAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACT 652
                  ||| ||||||||| |||  || |  |||  |  |||| | ||| |||
Db   98186 ------------AGGATGAAAAGGGTCTCAGAGCTACTTTTGTAGGTTAAACTGAAACT 98232

Qy     653 TCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAA 712
           |  | ||  || |  ||||| ||    | || |  | |||||||| |||||||| | |
Db   98233 TGTATACCATCACAATAGTCAAGTAGAGACCCAAGAGTAGGAAAGACTAGGATTGTAGAG 98292

Qy     713 TGCTGCA 719
             ||| |
Db   98293 CTCTGTA 98299
```

```
RESULT  26
BTAENSULF
LOCUS       BTAENSULF                231 bp    mRNA    linear   MAM 09-MAY-1998
DEFINITION  B.taurus mRNA for alpha endosulfine.
ACCESSION   X95958
VERSION     X95958.1  GI:1502303
KEYWORDS    alpha endosulfine.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 231)
  AUTHORS   Peyrollier,K., Heron,L., Virsolvy-Vergine,A., Le Cam,A. and
            Bataille,D.
  TITLE     Alpha endosulfine is a novel molecule, structurally related to a
            family of phosphoproteins
  JOURNAL   Biochem. Biophys. Res. Commun. 223 (3), 583-586 (1996)
  MEDLINE   96268722
REFERENCE   2  (bases 1 to 231)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-1996) D. Bataille, INSERM U 376, CHU
            Arnaud-de-Villeneuve, 371 Rue du Doyen Gaston Giraud, F- 34295
            Montpellier Cedex 5, FRANCE
COMMENT     Related sequence:
            Virsolvy-Vergine, Diabetologia 39, 135-141, 1996.
FEATURES             Location/Qualifiers
     source          1. .231
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /dev_stage="adult"
     CDS             <1. .>231
                     /codon_start=1
                     /product="alpha endosulfine"
                     /protein_id="CAA65196.1"
                     /db_xref="GI:1502304"
                     /db_xref="SPTREMBL:Q95105"
                     /translation="EGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSG
                     DYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQ"
BASE COUNT       81 a     58 c     57 g     35 t
ORIGIN


  Query Match             19.3%;  Score 213.4;  DB 4;  Length 231;
  Best Local Similarity   95.2%;  Pred. No. 1.2e-51;
  Matches  220;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy     179 GAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTA 238
           ||||||||||||||||||||||||| ||||||||||||| |||||||| |||||||||
Db       1 GAAGGTATTCTGCCTGAGAGAGCTGAGGAGGCAAAGCTAAAAGCCAAATATCCAAGCCTA 60

Qy     239 GGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTAC 298
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      61 GGACAAAAGCCTGGAGGTTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTAC 120

Qy     299 TTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGT 358
           |||||||||||||||||||||||||||||||||||||||||| |||||||| || |||
Db     121 TTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAACAAGCAGCTACCGAGT 180

Qy     359 GCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAG 409
           |||||||||||||||||||| |||||||| ||||| ||||| ||||||
Db     181 GCAGGACCAGACAAGAACCTGGTCACTGGTGACCACATTCCCACTCCACAG 231


RESULT  27
AC096386
LOCUS       AC096386              194799 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-39F15, WORKING DRAFT SEQUENCE, 26
            unordered pieces.
ACCESSION   AC096386
VERSION     AC096386.2  GI:17944296
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 194799)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 194799)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 20, 2001 this sequence version replaced gi:15628007.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GFEO
                Center clone name: CH230-39F15
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 183187 bases at least Q40
                Consensus quality: 186798 bases at least Q30
                Consensus quality: 189155 bases at least Q20
                Estimated insert size: 188406; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 4.4x in Q20 bases; sum-of-contigs estimation
            --------------
```

```
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 26 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1    36032: contig of 36032 bp in length
          *   36033    36132: gap of unknown length
          *   36133    47487: contig of 11355 bp in length
          *   47488    47587: gap of unknown length
          *   47588    62906: contig of 15319 bp in length
          *   62907    63006: gap of unknown length
          *   63007    77785: contig of 14779 bp in length
          *   77786    77885: gap of unknown length
          *   77886    84355: contig of 6470 bp in length
          *   84356    84455: gap of unknown length
          *   84456    97138: contig of 12683 bp in length
          *   97139    97238: gap of unknown length
          *   97239   110981: contig of 13743 bp in length
          *  110982   111081: gap of unknown length
          *  111082   122243: contig of 11162 bp in length
          *  122244   122343: gap of unknown length
          *  122344   131509: contig of 9166 bp in length
          *  131510   131609: gap of unknown length
          *  131610   139176: contig of 7567 bp in length
          *  139177   139276: gap of unknown length
          *  139277   146433: contig of 7157 bp in length
          *  146434   146533: gap of unknown length
          *  146534   152102: contig of 5569 bp in length
          *  152103   152202: gap of unknown length
          *  152203   159707: contig of 7505 bp in length
          *  159708   159807: gap of unknown length
          *  159808   166348: contig of 6541 bp in length
          *  166349   166448: gap of unknown length
          *  166449   171169: contig of 4721 bp in length
          *  171170   171269: gap of unknown length
          *  171270   174522: contig of 3253 bp in length
          *  174523   174622: gap of unknown length
          *  174623   177390: contig of 2768 bp in length
          *  177391   177490: gap of unknown length
          *  177491   180169: contig of 2679 bp in length
          *  180170   180269: gap of unknown length
          *  180270   183050: contig of 2781 bp in length
          *  183051   183150: gap of unknown length
          *  183151   184864: contig of 1714 bp in length
          *  184865   184964: gap of unknown length
          *  184965   187277: contig of 2313 bp in length
          *  187278   187377: gap of unknown length
          *  187378   188735: contig of 1358 bp in length
          *  188736   188835: gap of unknown length
          *  188836   190281: contig of 1446 bp in length
          *  190282   190381: gap of unknown length
          *  190382   191562: contig of 1181 bp in length
          *  191563   191662: gap of unknown length
          *  191663   193355: contig of 1693 bp in length
          *  193356   193455: gap of unknown length
          *  193456   194799: contig of 1344 bp in length.
FEATURES             Location/Qualifiers
     source          1. .194799
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-39F15"
BASE COUNT    54326 a  41772 c  42294 g  53891 t   2516 others
ORIGIN


  Query Match             17.0%;  Score 187.8;  DB 2;  Length 194799;
  Best Local Similarity   63.8%;  Pred. No. 2.1e-43;
  Matches  473;  Conservative    0;  Mismatches  207;  Indels   61;  Gaps   10;

Qy      240 GACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACT 299
            || || ||       || || ||||| ||||||  ||||||   ||||||||||||||||
Db     9918 GAGAACAGAGGCCCAGCACCAACTTCTTCATGACAAGACTC---AAAGGGCAAAAGTACT 9974

Qy      300 TTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTG 359
            |||    || | || |||   |  ||| |||||| ||| |     |||| |||| |  ||
Db     9975 TTG----AGAACACAACAGTGTCACCAGAGCCAACATGGA----CAGCAACTGCTATATG 10026

Qy      360 CAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCC 419
            |||| | || |||||| ||||||||  |||| ||||  ||  |||||||  || ||||||
Db     10027 CAGGGCTAGGCAAGAATCTGGTGACCAGTGACCACAT-GCCTTCCCACAGCATGTGCCCC 10085

Qy      420 AGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTC 479
            |||||   ||||| ||   || |||||||||  ||||  | |     | | ||||
Db     10086 AGAGAGTATCCTCACTTAACAGCAGCAAGCTCTCGGGTAACCTGTCTCTCTCCTC----- 10140

Qy      480 CCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTC 539
                  ||   | |   ||| | ||| | ||||| ||||| | |||||| || |||
Db     10141 ----------ACTCATTGGATTTTCTCTATCACAGGCAAGGATGGACTGGGCATCTTGTC 10190

Qy      540 AGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGG 599
            |||    ||||| || |||||| | ||   |  || |||| ||| ||| |||||| ||||
Db     10191 AGA---CCTCAGTTTACTAGCCTGGAAAAGCAATGCTTCTTCTGAGGGTTGCTGAAAAGG 10247

Qy      600 TAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACAC 659
             ||||| |||||| | |||||||| |||| || | ||| | ||  |||| | | ||||
Db     10248 CAATGTGGGTTGAGA-GGGGCTCTGAGTTCATGTCTTTAGATAACTTGAGATTCCCACGA 10306

Qy      660 ACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCA 719
             | | |||    |  ||  |  |||||||  || ||||  ||   |  | ||| ||||
Db     10307 TCCCACTGAGCCCAGGGAGGCACCTAGAAGGAGAAAAGTATAAACTGAGTTAAAGCTGAT 10366

Qy      720 AATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTC 779
            ||  |||| | | |   | |     ||                          || ||||||
Db     10367 AACTCTTTCTCTTTTGAACACTGGAGA----------------------AGTGAATGTT 10403

Qy      780 CCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAA 839
            ||| |||   || |||  |  ||  || ||| |  |||| ||||| ||  ||||| ||||
Db     10404 CCAGAATCAGTTGGGCACAATCTAAGG-ATAATACGGCAAGTTTAGCACACCAAGACTAA 10462

Qy      840 CCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTC 899
            || ||| | | |||||||||| ||| |||  | ||  |||||| | || ||||||   ||
Db     10463 CCCAACATGGCTGGGAAAGGCAGAGTGATGGTGATGTGTTTTTTATGTTTCTAATAATTT 10522

Qy      900 TGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAA 959
            | ||||||      | |||||| || |||||||| |   || ||||||||| ||| |||
Db     10523 TATCCCTT------CAATTGCTACAGGTTTCCATAAAAACATGTTTAGAGAACCCTCAAA 10576

Qy      960 AAGAAAAATTGACTTGCTTGC 980
            ||| |||||| |   | | ||
Db     10577 AAGGAAAATTTAAAGGGTAGC 10597


RESULT  28
MMA005983
LOCUS       MMA005983               339 bp    mRNA    linear   ROD 07-JAN-1999
DEFINITION  Mus musculus mRNA for cAMP-regulated phosphoprotein (ARPP-19).
ACCESSION   AJ005983
VERSION     AJ005983.1  GI:4138189
KEYWORDS    ARPP-19 gene; cAMP-regulated phosphoprotein.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Bataille D., U 376 'Endocrinologie des
            Peptides et regulation genique', INSERM, CHU Arnaud-de-Villeneuve,
            34295, FRANCE
REFERENCE   2  (bases 1 to 339)
  AUTHORS   Girault,J.A., Horiuchi,A., Gustafson,E., Rosen,N. and Greengard,P.
```

```
  TITLE     Differential expression of ARPP-16 and ARPP-19, two highly related
            phosphoproteins, one of which is specifically associated with
            dopamine-innervated brain regions
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /tissue_type="brain"
     gene            1. .339
                     /gene="ARPP-19"
     CDS             1. .339
                     /gene="ARPP-19"
                     /codon_start=1
                     /product="cAMP-regulated phosphoprotein"
                     /protein_id="CAA06797.1"
                     /db_xref="GI:4138190"
                     /translation="MSAEVPEAASAEEQKEMEDKVTSPEKAEEAKLKARYPHLGQKPG
                     GSDFLRKRLQKGQKYFDSGDYNMAKAKMKNKQLPAAAPDKTEVTGDHIPTPQDLPQRK
                     PSLVASKLAG"
BASE COUNT      111 a      73 c      91 g      64 t
ORIGIN
```

```
  Query Match             16.0%;  Score 176.2;  DB 10;  Length 339;
  Best Local Similarity   72.7%;  Pred. No. 1.2e-40;
  Matches  242;  Conservative    0;  Mismatches   88;  Indels    3;  Gaps    1;

Qy     128 GAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATT 187
           ||   ||| | ||  | | | |  |||||| |  |||| | | | || |||| |  || |
Db      10 GAAGTCCCCGAGGCAGCGTCTGCGGAGGAGCAAAAGGAAATGGAAGATAAAGTAACTAGT 69

Qy     188 CTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAG 247
              || |||| ||||||||| |||||| ||||||| |  || ||    | | |||||||||
Db      70 ---CCAGAGAAAGCTGAAGAAGCAAAGTTAAAGGCAAGGTATCCTCACTTGGGACAAAAG 126

Qy     248 CCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCA 307
           |||||||| || || ||  | | ||| ||| | ||||||||||||||||| ||||| ||
Db     127 CCTGGAGGGTCTGATTTTTTAAGGAAAAGATTGCAGAAAGGGCAAAAGTATTTTGATTCT 186

Qy     308 GGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCA 367
           || ||||||||||||| ||||| |||||||||| |||||||| ||   |||||  ||
Db     187 GGGGACTACAACATGGCAAAAGCAAAGATGAAGAACAAGCAGCTTCCTGCTGCAGCCCCG 246

Qy     368 GACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAG 427
           || ||||    ||| |||||||| ||||| ||||| |||||||| || || ||| | ||
Db     247 GATAAGACAGAGGTCACTGGTGACCACATTCCCACTCCACAGGACCTTCCTCAGCGGAAA 306

Qy     428 TCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAA 460
             | || || ||  | ||||| || || || | |
Db     307 CCATCCCTTGTTGCTAGCAAACTGGCTGGCTGA 339
```

```
RESULT  29
RNA005982
LOCUS       RNA005982               339 bp    mRNA    linear   ROD 07-JAN-1999
DEFINITION  Rattus norvegicus mRNA for cAMP-regulated phosphoprotein (ARPP-19).
ACCESSION   AJ005982
VERSION     AJ005982.1  GI:4138423
KEYWORDS    ARPP-19 gene; cAMP-regulated phosphoprotein.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Bataille D., U 376 'Endocrinologie des
            Peptides et regulation genique', INSERM, CHU Arnaud-de-Villeneuve,
            34295, FRANCE
REFERENCE   2  (bases 1 to 339)
  AUTHORS   Girault,J.A., Horiuchi,A., Gustafson,E., Rosen,N. and Greengard,P.
```

```
  TITLE     Differential expression of ARPP-16 and ARPP-19, two highly related
            cAMP-regulated phosphoproteins, one of which is specifically
            associated with dopamine-innervated brain regions
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /tissue_type="brain"
     gene            1. .339
                     /gene="ARPP-19"
     CDS             1. .339
                     /gene="ARPP-19"
                     /codon_start=1
                     /product="cAMP-regulated phosphoprotein"
                     /protein_id="CAA06796.1"
                     /db_xref="GI:4138424"
                     /translation="MSAEVPEAASAEEQKEMEDKVTSPEKAEEAKLKARYPHLGQKPG
                     GSDFLRKRLQKGQKYFDSGDYNMAKAKMKNKQLPAAAPDKTEVTGDHIPTPQDLPQRK
                     PSLVASKLAG"
BASE COUNT      112 a      72 c      89 g      66 t
ORIGIN
```

```
  Query Match             15.5%;  Score 171;  DB 10;  Length 339;
  Best Local Similarity   70.6%;  Pred. No. 4e-39;
  Matches  228;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy     138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
           | ||||   |    | |||| ||||  |       |||  |  |||  | |   || ||||
Db      17 CCGAGGCAGCGTCGGCGGAGGAGCAAAAAGAAATGGAAGATAAAGTAACTAGTCCAGAGA 76

Qy     198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
           |||||||| |||||| |||| || |  || ||     | ||||||||||||| || |
Db      77 AAGCTGAAGAAGCAAAGTTAAAAGCAAGGTATCCTCATTTGGGACAAAAGCCTGGTGGTT 136

Qy     258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
           | |||||  | | ||| ||| | ||||||||||||||||| ||||| || || || ||||
Db     137 CTGACTTTTTAAGGAAAAGATTGCAGAAAGGGCAAAAGTATTTTGATTCTGGGGATTACA 196

Qy     318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
           ||||||| ||||| |||||||||| |||||||| ||   |||||  || || ||||
Db     197 ACATGGCAAAAGCAAAGATGAAGAACAAGCAGCTTCCTGCTGCAGCCCCGGATAAGACAG 256

Qy     378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
           ||| |||||||| ||||| ||||||| || ||| | ||  | || || |
Db     257 AGGTCACTGGTGACCACATTCCCACTCCACAGGACCTTCCTCAGCGGAAACCATCCCTTG 316

Qy     438 TCACCAGCAAGCTTGCGGGGTAA 460
           |  | ||||| || || || | |
Db     317 TTGCTAGCAAACTGGCTGGCTGA 339
```

```
RESULT  30
HOS223091
LOCUS       HOS223091               339 bp    mRNA    linear   PRI 22-SEP-1998
DEFINITION  Homo sapiens mRNA for ARPP-19 protein.
ACCESSION   AJ223091
VERSION     AJ223091.1  GI:3647145
KEYWORDS    ARPP-19 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Heron,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-DEC-1997) Heron L., Endocrinologie des peptides,
            INSERM U 376, CHU Arnaud-de-Villeneuve, 34295 Montpellier, FRANCE
REFERENCE   2  (bases 1 to 339)
  AUTHORS   Heron,L., Virsolvy,A., Peyrollier,K., Gribble,F.M., Le Cam,A.,
            Ashcroft,F.M. and Bataille,D.
  TITLE     Human alpha-endosulfine, a possible regulator of
```

```
            sulfonylurea-sensitive KATP channel: molecular cloning, expression
            and biological properties
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (14), 8387-8391 (1998)
  MEDLINE   98318667
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="brain"
     gene            1. .339
                     /gene="ARPP-19"
     CDS             1. .339
                     /gene="ARPP-19"
                     /codon_start=1
                     /protein_id="CAA11115.1"
                     /db_xref="GI:3647146"
                     /translation="MSAEVPEAASAEEQKEMEDKVTSPEKAEEAKLKARYPHLGQKPG
                     GSDFLRKRLQKGQKYFDSGDYNMAKAKMKNKQLPTAAPDKTEVTGDHIPTPQDLPQRK
                     PSLVASKLAG"
BASE COUNT      112 a     76 c     87 g     64 t
ORIGIN


  Query Match             15.4%;  Score 169.4;  DB 9;  Length 339;
  Best Local Similarity   70.3%;  Pred. No. 1.2e-38;
  Matches  227;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy     138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
           | ||||   ||  || |||| ||||| |       |||  |  |||  | |   || ||||
Db      17 CCGAGGCAGCCTCCGCGGAGGAGCAGAAGGAAATGGAAGATAAAGTGACTAGTCCAGAGA 76

Qy     198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
            ||| ||||| |||||  |||| || | ||| ||    || ||||||||||||||||| |
Db      77 AAGCAGAAGAAGCAAAATTAAAAGCAAGATATCCTCATCTGGGACAAAAGCCTGGAGGTT 136

Qy     258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
           |||| ||| | | |||  |  | ||||||||||||||| || ||||| || || || ||||
Db     137 CCGATTTCTTAAGGAAACGGTTGCAGAAAGGGCAAAAATATTTTGATTCTGGGGATTACA 196

Qy     318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
           ||||||| ||||| || |||||||| ||||| || || | |||||  || || ||||
Db     197 ACATGGCTAAAGCAAAAATGAAGAACAAGCAACTTCCTACTGCAGCTCCGGATAAGACGG 256

Qy     378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
            ||| ||||||||| ||||| ||||| || || || || || ||  | ||| | || || |
Db     257 AGGTCACTGGTGACCACATTCCCACTCCGCAAGACCTTCCTCAACGGAAGCCGTCCCTTG 316

Qy     438 TCACCAGCAAGCTTGCGGGGTAA 460
           |  | ||||| || || || | |
Db     317 TTGCTAGCAAACTGGCTGGCTGA 339


RESULT  31
BOVARPP19
LOCUS       BOVARPP19               1188 bp    mRNA    linear   MAM 26-APR-1993
DEFINITION  Bovine cAMP-regulated phosphoprotein (ARPP-19) mRNA, complete cds.
ACCESSION   M33618 J05450
VERSION     M33618.1  GI:162689
KEYWORDS    ARPP-19; cAMP-regulated phosphoprotein.
SOURCE      Bovine young adult caudate nuclei, cDNA to mRNA, clone 1.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1188)
  AUTHORS   Horiuchi,A., Williams,K.R., Kurihara,T., Nairn,A.C. and
            Greengard,P.
  TITLE     Purification and cDNA cloning of ARPP-16, a cAMP-regulated
            phosphoprotein enriched in basal ganglia, and of a related
            phosphoprotein, ARPP-19
  JOURNAL   J. Biol. Chem. 265, 9476-9484 (1990)
  MEDLINE   90264450
COMMENT     Draft entry and computer-readable sequence for [1] kindly submitted
            by A.Horiuchi, 06-APR-1990.
FEATURES             Location/Qualifiers
     source          1. .1188
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
     mRNA            1. .1188
                     /product="ARPP-19 mRNA"
     CDS             97. .435
                     /note="cAMP-regulated phosphoprotein"
                     /codon_start=1
                     /protein_id="AAA30386.1"
                     /db_xref="GI:162690"
                     /translation="MSAEVPEAASAEEQKEMEDKVTSPEKAEEAKLKARYPHLGQKPG
                     GSDFLRKRLQKGQKYFDSGDYNMAKAKMKNKQLPTATPDKTEVTGDHIPTPQDLPQRK
                     PSLVASKLAG"
BASE COUNT      349 a     211 c     261 g     367 t
ORIGIN


  Query Match             15.4%;  Score 169.4;  DB 4;  Length 1188;
  Best Local Similarity   70.3%;  Pred. No. 1.6e-38;
  Matches  227;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy     138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
           | ||||   ||  || |||| ||||| |       |||  |  |||  | |   || ||||
Db     113 CCGAGGCAGCCTCCGCGGAGGAGCAGAAGGAAATGGAAGATAAAGTGACTAGTCCAGAGA 172

Qy     198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
            ||||||||| |||||  |||| || |  || ||    || ||||||||||||||||| |
Db     173 AAGCTGAAGAAGCAAAATTAAAAGCAAGGTATCCTCATCTGGGACAAAAGCCTGGAGGTT 232

Qy     258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
           |||| ||  | | |||  || | ||||||||||||| || ||||| || || || ||||
Db     233 CCGATTTTTTAAGGAAACGATTGCAGAAAGGGCAAAAATATTTTGATTCTGGGGATTACA 292

Qy     318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
           |||||||  ||||| || |||||||| ||||| || || | ||||   || || ||||
Db     293 ACATGGCTAAAGCAAAAATGAAGAACAAGCAACTTCCTACTGCAACCCCGGATAAGACAG 352

Qy     378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
            ||| ||||||||| ||||| ||||| |||||||| || || ||  | ||  | || || |
Db     353 AGGTCACTGGTGACCACATTCCCACTCCACAGGACCTTCCTCAACGGAAACCATCTCTTG 412

Qy     438 TCACCAGCAAGCTTGCGGGGTAA 460
           |  | ||||| || || || | |
Db     413 TTGCTAGCAAACTGGCTGGCTGA 435


RESULT  32
BC003418
LOCUS       BC003418                1228 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, cyclic AMP phosphoprotein, 19 kD, clone MGC:5468
            IMAGE:3451558, mRNA, complete cds.
ACCESSION   BC003418
VERSION     BC003418.1  GI:13097335
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```

```
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalon@bcm.tmc.edu.
            Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
            A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
            Muzny,D.M., Gibbs,R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 3 Row: p Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Similarity but not identity to protein.
FEATURES              Location/Qualifiers
     source           1. .1228
                      /organism="Homo sapiens"
                      /db_xref="LocusID:10776"
                      /db_xref="taxon:9606"
                      /clone="MGC:5468 IMAGE:3451558"
                      /tissue_type="Placenta, choriocarcinoma"
                      /clone_lib="NIH_MGC_10"
                      /lab_host="DH10B"
                      /note="Vector: pCMV-SPORT6"
     CDS              125. .463
                      /codon_start=1
                      /product="cyclic AMP phosphoprotein, 19 kD"
                      /protein_id="AAH03418.1"
                      /db_xref="GI:13097336"
                      /translation="MSAEVPEAASAEEQKEMEDKVTSPEKAEEAKLKARYPHLGQKPG
                      GSDFLRKRLQKGQKYFDSGDYNMAKAKMKNKQLPTAAPDKTEVTGDHIPTPQDLPQRK
                      PSLVASKLAG"
BASE COUNT      354 a    229 c    272 g    373 t
ORIGIN


  Query Match             15.4%;  Score 169.4;  DB 9;  Length 1228;
  Best Local Similarity   70.3%;  Pred. No. 1.6e-38;
  Matches  227;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy    138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
          | ||||   ||  || |||| ||||| |      ||| |  ||| | | |   || ||||
Db    141 CCGAGGCAGCCTCCGCGGAGGAGCAGAAGGAAATGGAAGATAAAGTGACTAGTCCAGAGA 200

Qy    198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
           ||| ||||| ||||| |||| || | ||| ||    || ||||||||||||||||| |
Db    201 AAGCAGAAGAAGCAAAATTAAAAGCAAGATATCCTCATCTGGGACAAAAGCCTGGAGGTT 260

Qy    258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
          | || ||| | | |||  |  | ||||||||||||||| || ||||| || || || ||||
Db    261 CAGATTTCTTAAGGAAACGGTTGCAGAAAGGGCAAAAATATTTTGATTCTGGGGATTACA 320

Qy    318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
          ||||||| ||||| || |||||||| ||||| || || | |||||  || || ||||
Db    321 ACATGGCTAAAGCAAAAATGAAGAACAAGCAACTTCCTACTGCAGCTCCGGATAAGACGG 380

Qy    378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
           ||| |||||||| ||||| ||||| || || || || || ||  | ||| | || || |
Db    381 AGGTCACTGGTGACCACATTCCCACTCCGCAAGACCTTCCTCAACGGAAGCCGTCCCTTG 440

Qy    438 TCACCAGCAAGCTTGCGGGGTAA 460
          |  | |||||||| || || | |
Db    441 TTGCTAGCAAGCTGGCTGGCTGA 463


RESULT  33
AF084555
LOCUS       AF084555               5171 bp    mRNA    linear   PRI 16-AUG-2001
DEFINITION  Homo sapiens okadaic acid-inducible and cAMP-regulated
            phosphoprotein 19 (ARPP-19) mRNA, complete cds.
ACCESSION   AF084555
VERSION     AF084555.1  GI:5813858
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5171)
  AUTHORS   Chin,L.S., Singh,S.K., Wang,Q. and Murray,S.F.
  TITLE     Identification of okadaic-acid-induced genes by mRNA differential
            display in glioma cells
  JOURNAL   J. Biomed. Sci. 7 (2), 152-159 (2000)
  MEDLINE   20219739
   PUBMED   10754390
REFERENCE   2  (bases 1 to 5171)
  AUTHORS   Singh,S.K., Murray,S.F. and Chin,L.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-1998) Neurosurgery, University of Maryland at
            Baltimore, 685 W. Baltimore Street, Baltimore, MD 21201, USA
FEATURES              Location/Qualifiers
     source           1. .5171
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="OA4-5"
                      /tissue_type="brain"
                      /dev_stage="fetus"
     gene             1. .5171
                      /gene="ARPP-19"
     CDS              125. .463
                      /gene="ARPP-19"
                      /codon_start=1
                      /product="okadaic acid-inducible and cAMP-regulated
                      phosphoprotein 19"
                      /protein_id="AAD52044.1"
                      /db_xref="GI:5813859"
                      /translation="MSAEVPEAASAEEQKEMEDKVTSPEKAEEAKLKARYPHLGQKPG
                      GSDFLRKRLQKGQKYFDSGDYNMAKAKMKNKQLPTAAPDKTEVTGDHIPTPQDLPQRK
                      PSLVASKLAG"
BASE COUNT     1494 a    974 c    956 g   1746 t      1 others
ORIGIN


  Query Match             15.4%;  Score 169.4;  DB 9;  Length 5171;
  Best Local Similarity   70.3%;  Pred. No. 2.3e-38;
  Matches  227;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy    138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
          | ||||   ||  || |||| ||||| |      ||| |  ||| | | |   || ||||
Db    141 CCGAGGCAGCCTCCGCGGAGGAGCAGAAGGAAATGGAAGATAAAGTGACTAGTCCAGAGA 200

Qy    198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
           ||| ||||| ||||| |||| || | ||| ||    || ||||||||||||||||| |
Db    201 AAGCAGAAGAAGCAAAATTAAAAGCAAGATATCCTCATCTGGGACAAAAGCCTGGAGGTT 260

Qy    258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
          | || ||| | | |||  |  | ||||||||||||||| || ||||| || || || ||||
Db    261 CAGATTTCTTAAGGAAACGGTTGCAGAAAGGGCAAAAATATTTTGATTCTGGGGATTACA 320

Qy    318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
          ||||||| ||||| || |||||||| ||||| || || | |||||  || || ||||
Db    321 ACATGGCTAAAGCAAAAATGAAGAACAAGCAACTTCCTACTGCAGCTCCGGATAAGACGG 380

Qy    378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
           ||| |||||||| ||||| ||||| || || || || || ||  | ||| | || || |
Db    381 AGGTCACTGGTGACCACATTCCCACTCCGCAAGACCTTCCTCAACGGAAGCCGTCCCTTG 440

Qy    438 TCACCAGCAAGCTTGCGGGGTAA 460
          |  | |||||||| || || | |
Db    441 TTGCTAGCAAGCTGGCTGGCTGA 463


RESULT  34
SSA005981
LOCUS       SSA005981              339 bp    mRNA    linear   MAM 07-JAN-1999
DEFINITION  Sus scrofa mRNA for cAMP-regulated phosphoprotein (ARPP-19).
```

```
ACCESSION   AJ005981
VERSION     AJ005981.1  GI:4138554
KEYWORDS    ARPP-19 gene; cAMP-regulated phosphoprotein.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Bataille,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Bataille D., U 376 'Endocrinologie des
            Peptides et regulation genique', INSERM, CHU Arnaud-de-Villeneuve,
            34295, FRANCE
REFERENCE   2  (bases 1 to 339)
  AUTHORS   Girault,J.A., Horiuchi,A., Gustafson,E.L., Rosen,N.L. and
            Greengard,P.
  TITLE     Differential expression of ARPP-16 and ARPP-19, two highly related
            cAMP-regulated phosphoproteins, one of which is specifically
            associated with dopamine-innervated brain regions
  JOURNAL   J. Neurosci. 10 (4), 1124-1133 (1990)
  MEDLINE   90229913
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
     gene            1. .339
                     /gene="ARPP-19"
     CDS             1. .339
                     /gene="ARPP-19"
                     /codon_start=1
                     /product="cAMP-regulated phosphoprotein"
                     /protein_id="CAA06795.1"
                     /db_xref="GI:4138555"
                     /translation="MSAEVPEAASAEEQKEMEDKVTSPEKAEEAKLKARYPHLGQKPG
                     GSDFLRKRLQKGQKYFDSGDYNMAKAKMKNKQLPTAAPDKTEVTGDHIPTPQDLPQRK
                     PSLVASKLAG"
BASE COUNT      115 a     74 c     84 g     66 t
ORIGIN


  Query Match           15.2%;  Score 167.8;  DB 4;  Length 339;
  Best Local Similarity   70.0%;  Pred. No. 3.5e-38;
  Matches  226;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy    138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
          | ||||   ||  || |||| ||||| |      ||| | ||| | | |  || ||||
Db     17 CCGAGGCAGCCTCCGCGGAGGAGCAGAAGGAAATGGAAGATAAAGTGACTAGTCCAGAGA 76

Qy    198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
          |||||||| ||||| |||| || |  || ||    || ||||||||||||||| |
Db     77 AAGCTGAAGAAGCAAAATTAAAAGCGAGGTATCCTCATCTGGGACAAAAGCCTGGAGGTT 136

Qy    258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
          | || ||| | |  ||  || | |||||||||||||| || ||||| || || || || |
Db    137 CAGATTTCTTAAGAAAACGATTGCAGAAAGGGCAAAAATATTTTGATTCTGGGGATTATA 196

Qy    318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
          ||||||| ||||| || |||||||| ||||| || || | |||||  || || ||||
Db    197 ACATGGCTAAAGCAAAAATGAAGAACAAGCAACTTCCTACTGCAGCCCCGGATAAGACAG 256

Qy    378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
          ||| |||||||| ||||| ||||| |||||||| || || ||  | ||  | || || |
Db    257 AGGTCACTGGTGACCACATTCCCACTCCACAGGACCTTCCTCAACGGAAACCATCTCTTG 316

Qy    438 TCACCAGCAAGCTTGCGGGGTAA 460
          |  | ||||| || || || | |
Db    317 TTGCTAGCAAACTGGCTGGCTGA 339


RESULT  35
AC091874
LOCUS       AC091874               71941 bp    DNA     linear   HTG 09-JUN-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2593A12, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC091874
VERSION     AC091874.1  GI:14333810
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 71941)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71941)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 825705
            Center clone name: CITB-E1_2593A12
            -------
            Summary Statistics
            Consensus quality: 62238 bases at least Q40
            Consensus quality: 67347 bases at least Q30
            Consensus quality: 68096 bases at least Q20
            Estimated insert size: 219000; agarose-fp estimation
            Estimated insert size: 71141; sum-of-contigs estimation
            Quality coverage:  4.14 in Q20 bases; agarose-fp estimation
            Quality coverage:  12.76 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1461: contig of 1461 bp in length
            *     1462     1561: gap of unknown length
            *     1562     2633: contig of 1072 bp in length
            *     2634     2733: gap of unknown length
            *     2734     4507: contig of 1774 bp in length
            *     4508     4607: gap of unknown length
            *     4608     7249: contig of 2642 bp in length
            *     7250     7349: gap of unknown length
            *     7350    12208: contig of 4859 bp in length
            *    12209    12308: gap of unknown length
            *    12309    20679: contig of 8371 bp in length
            *    20680    20779: gap of unknown length
            *    20780    29687: contig of 8908 bp in length
            *    29688    29787: gap of unknown length
            *    29788    39781: contig of 9994 bp in length
            *    39782    39881: gap of unknown length
            *    39882    71941: contig of 32060 bp in length.
FEATURES             Location/Qualifiers
     source          1. .71941
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2593A12"
                     /clone_lib="CalTech human BAC library D"
BASE COUNT    20351 a  15896 c  16000 g  18892 t    802 others
ORIGIN


  Query Match           15.1%;  Score 166.8;  DB 2;  Length 71941;
  Best Local Similarity   69.9%;  Pred. No. 2.6e-37;
  Matches  225;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;
```

```
Qy     139 GGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAG 198
           |||||   ||  || |||| ||||| |     ||||  |  |||  | |   || ||||
Db   16537 GGAGGCAGCCTCCGCGGAGGAGCAGAAGGAAAAGGAAGATAAAGTGACTAGTCCAGAGAA 16596

Qy     199 AGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTC 258
           ||||||||| |||||  |||| || |  || ||    || ||||||||||||||||| ||
Db   16597 AGCTGAAGAAGCAAAATTAAAAGCAAGGTATCCTCATCTGGGACAAAAGCCTGGAGGTTC 16656

Qy     259 CGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAA 318
           || ||| | | |||  || | ||||||||||||||| || ||||| || ||  | |||||
Db   16657 AGATTTCTTAAGGAAACGATTGCAGAAAGGGCAAAAATATTTTGATTCTGGGCATTACAA 16716

Qy     319 CATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCT 378
           || ||| ||||| || | |||||  ||||| || || | |||||  ||||| ||||
Db   16717 CAGGGCTAAAGCAAAAACGAAGAGCAAGCAACTTCCTACTGCAGCTCCAGATAAGACGGA 16776

Qy     379 GGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGT 438
           ||| |||||||| ||||| || || |||||||| || || ||  | ||  | || || ||
Db   16777 GGTCACTGGTGACCACATTCCTACTCCACAGGACCTTCCTCAACGGAAACCATCCCTTGT 16836

Qy     439 CACCAGCAAGCTTGCGGGGTAA 460
           | | |||||| || || | |
Db   16837 TGCTAACAAGCTGGCTGGCTGA 16858
```

RESULT 36
AK057673
LOCUS AK057673 2288 bp mRNA linear PRI 31-OCT-2001
DEFINITION Homo sapiens cDNA FLJ33111 fis, clone TRACH2001085.
ACCESSION AK057673
VERSION AK057673.1 GI:16553438
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens trachea cDNA to mRNA, clone_lib:TRACH2 clone:TRACH2001085.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K., Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2288)
AUTHORS Isogai,T., Otsuki,T. and Sugiyama,T.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

```
FEATURES             Location/Qualifiers
     source          1. .2288
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="TRACH2001085"
                     /tissue_type="trachea"
                     /clone_lib="TRACH2"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      604 a    462 c    597 g    625 t
ORIGIN
```

```
  Query Match             15.0%;  Score 165.4;  DB 9;  Length 2288;
  Best Local Similarity   96.6%;  Pred. No. 2.9e-37;
  Matches  169;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 81

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 141

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAG 175
           |||||||||||||||||||||||||||||||||||||||||||  | |  | ||
Db     142 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGTGAAGTGGAAG 196
```

RESULT 37
BOVARPP16
LOCUS BOVARPP16 1961 bp mRNA linear MAM 26-APR-1993
DEFINITION Bovine cAMP-regulated phosphoprotein (ARPP-16) mRNA, complete cds.
ACCESSION M33617 J05450
VERSION M33617.1 GI:162687
KEYWORDS ARPP-16; cAMP-regulated phosphoprotein.
SOURCE Bovine young adult caudate nuclei, cDNA to mRNA, clone 6.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 1961)
AUTHORS Horiuchi,A., Williams,K.R., Kurihara,T., Nairn,A.C. and Greengard,P.
TITLE Purification and cDNA cloning of ARPP-16, a cAMP-regulated phosphoprotein enriched in basal ganglia, and of a related phosphoprotein, ARPP-19
JOURNAL J. Biol. Chem. 265, 9476-9484 (1990)
MEDLINE 90264450
COMMENT Draft entry and computer-readable sequence for [1] kindly submitted by A.Horiuchi, 06-APR-1990.

```
FEATURES             Location/Qualifiers
     source          1. .1961
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
     mRNA            1. .1961
                     /product="ARPP-16 mRNA"
     misc_feature    85. .88
                     /note="phosphorylation site"
     CDS             112. .402
                     /note="cAMP-regulated phosphoprotein (ARPP-19)"
                     /codon_start=1
                     /protein_id="AAA30385.1"
                     /db_xref="GI:162688"
                     /translation="MEDKVTSPEKAEEAKLKARYPHLGQKPGGSDFLRKRLQKGQKYF
                     DSGDYNMAKAKMKNKQLPTATPDKTEVTGDHIPTPQDLPQRKPSLVASKLAG"
BASE COUNT      606 a    320 c    355 g    680 t
ORIGIN
```

```
  Query Match             14.5%;  Score 159.4;  DB 4;  Length 1961;
  Best Local Similarity   72.0%;  Pred. No. 1.6e-35;
  Matches  208;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy     172 GGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCC 231
           |||  |  |||  | |   || |||| ||||||||| |||||  |||| || |  || ||
Db     114 GGAAGATAAAGTGACTAGTCCAGAGAAAGCTGAAGAAGCAAAATTAAAAGCAAGGTATCC 173

Qy     232 AAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCA 291
             || |||||||||||||||| ||||| ||  | | |||  || | ||||||||||
Db     174 TCATCTGGGACAAAAGCCTGGAGGTTCCGATTTTTTAAGGAAACGATTGCAGAAAGGGCA 233

Qy     292 AAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCT 351
```

```
          ||| || ||||| || || || |||||||||||| ||||| || |||||||| ||||| ||
Db    234 AAAATATTTTGATTCTGGGGATTACAACATGGCTAAAGCAAAAATGAAGAACAAGCAACT 293

Qy    352 GCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGA 411
          || | ||||   || || ||||    ||| |||||||| ||||| ||||| ||||||||
Db    294 TCCTACTGCAACCCCGGATAAGACAGAGGTCACTGGTGACCACATTCCCACTCCACAGGA 353

Qy    412 TCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAA 460
          || || ||  | ||  | || || ||  | ||||| || || || | |
Db    354 CCTTCCTCAACGGAAACCATCTCTTGTTGCTAGCAAACTGGCTGGCTGA 402
```

RESULT 38
AC068538
LOCUS AC068538 173169 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-141G4 from 2, complete sequence.
ACCESSION AC068538
VERSION AC068538.8 GI:16931042
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 173169)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 173169)
AUTHORS VanBrunt,A., Haglund,K. and Doebber,A.
TITLE The sequence of Homo sapiens BAC clone RP11-141G4
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 173169)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-MAY-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 173169)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 173169)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 173169)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Nov 15, 2001 this sequence version replaced gi:16418235.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_NH0141G04
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-292K15; the clone sequenced to the right is RP11-480J15. Actual start of this clone is at base position 1 of RP11-141G4; actual end is at base position 173169 of RP11-141G4.

Data from AC092845 was used to finish this clone, AC068538. Polymorphisms have been identified between AC092845 and AC068538.

FEATURES Location/Qualifiers

source	1. .173169
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-141G4"
	/clone_lib="RPCI-11"
repeat_region	2. .803
	/rpt_family="ERV1"
repeat_region	812. .909
	/rpt_family="ERV1"
repeat_region	910. .1279
	/rpt_family="ERV1"
repeat_region	2183. .2217
	/rpt_family="ERV1"
repeat_region	2220. .2653
	/rpt_family="ERV1"
repeat_region	2654. .3050
	/rpt_family="MaLR"
repeat_region	3057. .3166
	/rpt_family="ERV1"
repeat_region	3639. .3981
	/rpt_family="L1"
repeat_region	4152. .4444
	/rpt_family="Alu"
repeat_region	4607. .5058
	/rpt_family="MaLR"
repeat_region	5362. .5449
	/rpt_family="MaLR"
repeat_region	5552. .5840
	/rpt_family="Alu"
repeat_region	5869. .5968
	/rpt_family="L1"
repeat_region	5955. .6018
	/rpt_family="CT-rich"
repeat_region	6379. .6783
	/rpt_family="ERVL"
repeat_region	6782. .6802
	/rpt_family="AT_rich"
repeat_region	6806. .6837
	/rpt_family="(TG)n"
repeat_region	6899. .7204
	/rpt_family="Alu"

```
repeat_region   7208. .7716
                /rpt_family="MER2_type"
repeat_region   7716. .7807
                /rpt_family="MER2_type"
repeat_region   8105. .8353
                /rpt_family="L1"
repeat_region   8610. .8980
                /rpt_family="MaLR"
repeat_region   9039. .9059
                /rpt_family="AT_rich"
repeat_region   9782. .10081
                /rpt_family="Alu"
repeat_region   10260. .10487
                /rpt_family="MIR"
repeat_region   12061. .12147
                /rpt_family="CR1"
repeat_region   12403. .12646
                /rpt_family="CR1"
repeat_region   12645. .12983
                /rpt_family="CR1"
repeat_region   13017. .13087
                /rpt_family="CR1"
repeat_region   13831. .14143
                /rpt_family="L1"
repeat_region   14231. .14268
                /rpt_family="(TA)n"
repeat_region   14583. .14867
                /rpt_family="Alu"
repeat_region   15161. .15391
                /rpt_family="MIR"
repeat_region   15906. .16029
                /rpt_family="MIR"
repeat_region   16219. .16364
                /rpt_family="L2"
repeat_region   16401. .16709
                /rpt_family="Alu"
repeat_region   16683. .16714
                /rpt_family="(GAAAA)n"
repeat_region   18661. .18943
                /rpt_family="Alu"
repeat_region   19142. .19417
                /rpt_family="Alu"
repeat_region   20259. .20316
                /rpt_family="MIR"
repeat_region   20650. .20746
                /rpt_family="AcHobo"
repeat_region   21564. .21733
                /rpt_family="MER1_type"
repeat_region   22242. .22348
                /rpt_family="MIR"
repeat_region   22408. .22556
                /rpt_family="MER103"
repeat_region   24087. .24364
                /rpt_family="ERV1"
repeat_region   24366. .24949
                /rpt_family="ERV1"
repeat_region   24950. .25377
                /rpt_family="ERV1"
repeat_region   25576. .25811
                /rpt_family="MIR"
repeat_region   25935. .26516
                /rpt_family="ERV1"
repeat_region   26517. .26825
                /rpt_family="Alu"
repeat_region   26826. .26973
                /rpt_family="ERV1"
repeat_region   28124. .28315
                /rpt_family="MIR"

Query Match             14.2%;  Score 156.6;  DB 9;  Length 173169;
Best Local Similarity   67.8%;  Pred. No. 3.2e-34;
Matches  219;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy     138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
           | ||||   ||   | |||| ||||| |      ||| |   ||  | |   || ||||
Db   49976 CCGAGGCAGCCTCTGTGGAGGAGCAGAAGGAAATGGAAGATACAGTGACTAGTCCAGAGA 50035

Qy     198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
           || |||||  |||||  |||| || | ||| ||    || ||||||| ||||||||| |
Db   50036 AAGTTGAAGCAGCAAAATTAAAAGCAAGATATCCTCATCTGGGACAAATGCCTGGAGGTT 50095

Qy     258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
           | || ||| | | |||  || | ||||||||||||||| || ||||| || || || ||||
Db   50096 CAGATTTCTTAAGGAAACGATTGCAGAAAGGGCAAAAATATTTTGATTCTGGGGATTACA 50155

Qy     318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
           ||||||| ||||| || |||||||  |||| || || | |||||  |  || ||||
Db   50156 ACATGGCTAAAGCAAAAATGAAGAACGAGCAACTTCCTACTGCAGCTCTGGATAAGATGG 50215

Qy     378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
           ||| || ||||| ||||| ||||| ||| |||| || || ||    ||  | ||  | |
Db   50216 AGGTCACGGGTGACCACATTCCCACTCCAGAGGACCTTCCTCAACAGAAACCATCCATTG 50275

Qy     438 TCACCAGCAAGCTTGCGGGGTAA 460
           |  | ||||||| || || | |
Db   50276 TTGCTAGCAAGCTGGCTGGTTGA 50298


RESULT  39
AC079548
LOCUS       AC079548              221899 bp    DNA     linear   HTG 02-SEP-2000
DEFINITION  Mus musculus clone RP23-416M3, WORKING DRAFT SEQUENCE, 19 unordered
            pieces.
ACCESSION   AC079548
VERSION     AC079548.1  GI:9964913
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 221899)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 221899)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 1896907
            Center clone name: RPCI-23_416M3
            -------
            Summary Statistics
            Consensus quality: 206046 bases at least Q40
            Consensus quality: 212579 bases at least Q30
            Consensus quality: 214022 bases at least Q20
            Estimated insert size: 179000; agarose-fp estimation
            Estimated insert size: 220099; sum-of-contigs estimation
            Quality coverage:  10.88 in Q20 bases; agarose-fp estimation
            Quality coverage:  8.84 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1178: contig of 1178 bp in length
            *     1179     1278: gap of unknown length
```

```
            *     1279     2327: contig of 1049 bp in length
            *     2328     2427: gap of unknown length
            *     2428     3561: contig of 1134 bp in length
            *     3562     3661: gap of unknown length
            *     3662     4786: contig of 1125 bp in length
            *     4787     4886: gap of unknown length
            *     4887     6634: contig of 1748 bp in length
            *     6635     6734: gap of unknown length
            *     6735     8429: contig of 1695 bp in length
            *     8430     8529: gap of unknown length
            *     8530     9649: contig of 1120 bp in length
            *     9650     9749: gap of unknown length
            *     9750    11496: contig of 1747 bp in length
            *    11497    11596: gap of unknown length
            *    11597    12762: contig of 1166 bp in length
            *    12763    12862: gap of unknown length
            *    12863    14663: contig of 1801 bp in length
            *    14664    14763: gap of unknown length
            *    14764    16240: contig of 1477 bp in length
            *    16241    16340: gap of unknown length
            *    16341    21281: contig of 4941 bp in length
            *    21282    21381: gap of unknown length
            *    21382    24293: contig of 2912 bp in length
            *    24294    24393: gap of unknown length
            *    24394    25587: contig of 1194 bp in length
            *    25588    25687: gap of unknown length
            *    25688    40191: contig of 14504 bp in length
            *    40192    40291: gap of unknown length
            *    40292    58318: contig of 18027 bp in length
            *    58319    58418: gap of unknown length
            *    58419    91749: contig of 33331 bp in length
            *    91750    91849: gap of unknown length
            *    91850   132661: contig of 40812 bp in length
            *   132662   132761: gap of unknown length
            *   132762   221899: contig of 89138 bp in length.
FEATURES             Location/Qualifiers
     source          1. .221899
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-416M3"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    62150 a  48312 c  48082 g  61476 t   1879 others
ORIGIN


  Query Match           10.3%;  Score 113.6;  DB 2;  Length 221899;
  Best Local Similarity   58.3%;  Pred. No. 1.6e-21;
  Matches  399;  Conservative    0;  Mismatches  224;  Indels   61;  Gaps    9;

Qy      45 TCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCG 104
           |||| || ||       | |  |  |   ||||  | ||  ||     |||  ||||||
Db   29971 TCCCTCTCCCAGCACAAGATGACACTTTGGGGCTGCCTATTCTGTGTGGTCTGGGTCCCA 30030

Qy     105 CCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGG 164
           ||||||||| ||| |||||||| |||  |||||  ||||| || | | ||||||  ||
Db   30031 CCATGTCCCTGAATCAAGAAGAGGAGGCCCCTGTAGAGGA-ACTGTCCAGGAGACACAAA 30089

Qy     165 ACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCA 224
           | ||  ||||||||||   |                                |  | |
Db   30090 ATACAGAGGAGAAAGACTTTCCAGG----------------------------AGAGTCT 30121

Qy     225 AATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGA 284
           | |   |||   |||||||| ||||    ||||| ||||  |||||  ||||||   |
Db   30122 CAAAGGTAAGGAGAGGACAAAGGCCT---AGCTCCAACTTCTCCATGACAAGACTC---A 30175

Qy     285 AAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATA 344
           |||| ||||||||||||||| |||  ||  ||   |  | |  ||||||||||
Db   30176 AAGGCCAAAAGTACTTTGACTGAGGGCACAGCAGTCTCACGAGTGCCAAGATGAAC---- 30231

Qy     345 AGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCC---CCA 401
           |||||||||  |  |||||||| ||||||||  |||||||||| | | ||| |   ||
Db   30232 AGCAGCTGCTATATGCAGGACC-GACAAGAAGTTGGTGACTGGTAACCCCATGCATGCCT 30290

Qy     402 CCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAAC 461
           |||||||  | |||||  ||||| |||  || |||||||||  | ||  | |
Db   30291 TCCCACAGCGTGTGCCCTAGAGAATATCCTCACTCAACAGCAGCAAGCTCTCAGGTAACC 30350

Qy     462 CTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGA 521
           |                | ||  ||  ||   |     | | | ||| || |||| | |
Db   30351 TGG--------------CTCTCTCTTCACTCATTAGATTCTCTGTATCATGGGCAAGCA 30395

Qy     522 TGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGC 581
           || | |  ||| || | | ||    ||||| || ||||||   ||  ||  ||  ||| |
Db   30396 TGGACTAAGCATCTTGCCGGA---CCTCAGTTTACTAGCCCAGAAAAACAATGCATCTTC 30452

Qy     582 TGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTT 641
           |||||| |       | |    |  || |  | |||  |||| |||| |||| ||| |||
Db   30453 TGGGGGGTTGCTGAGAAGATGAGGTGGGTTGAGAGGAACTCTGAGTTCATTTCTTTAGTT 30512

Qy     642 AGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTA 701
           || |||| | | |||||  | |  ||    |       |  |||||| |||| ||||  ||
Db   30513 AGCTTGAGATTCCCACAATCCCATTGGGTCCAGCAAGGCACCTAGACATAGAAAAGTATA 30572

Qy     702 GGATTGGATAATGCTGCAAATGCT 725
             |     ||   || | | | |
Db   30573 AATTGAATAAACCATGTAGAAGGT 30596


RESULT  40
AC090836/c
LOCUS       AC090836               70244 bp    DNA     linear   HTG 10-MAR-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-23N2 map 15, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC090836
VERSION     AC090836.1  GI:13270705
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 70244)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-23N2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 70244)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
```

```
    Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
    Center project name: L13121
    Center clone name: 23_N_2
--------------.
* NOTE: This record contains 87 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*       1      703: contig of 703 bp in length
*     704  803: gap of      100 bp
*     804     1525: contig of 722 bp in length
*    1526 1625: gap of      100 bp
*    1626     2347: contig of 722 bp in length
*    2348 2447: gap of      100 bp
*    2448     3133: contig of 686 bp in length
*    3134 3233: gap of      100 bp
*    3234     3943: contig of 710 bp in length
*    3944 4043: gap of      100 bp
*    4044     4754: contig of 711 bp in length
*    4755 4854: gap of      100 bp
*    4855     5563: contig of 709 bp in length
*    5564 5663: gap of      100 bp
*    5664     6380: contig of 717 bp in length
*    6381 6480: gap of      100 bp
*    6481     7207: contig of 727 bp in length
*    7208 7307: gap of      100 bp
*    7308     8022: contig of 715 bp in length
*    8023 8122: gap of      100 bp
*    8123     8826: contig of 704 bp in length
*    8827 8926: gap of      100 bp
*    8927     9602: contig of 676 bp in length
*    9603 9702: gap of      100 bp
*    9703    10410: contig of 708 bp in length
*   10411 10510: gap of      100 bp
*   10511    11209: contig of 699 bp in length
*   11210 11309: gap of      100 bp
*   11310    12023: contig of 714 bp in length
*   12024 12123: gap of      100 bp
*   12124    12827: contig of 704 bp in length
*   12828 12927: gap of      100 bp
*   12928    13638: contig of 711 bp in length
*   13639 13738: gap of      100 bp
*   13739    14445: contig of 707 bp in length
*   14446 14545: gap of      100 bp
*   14546    15212: contig of 667 bp in length
*   15213 15312: gap of      100 bp
*   15313    16027: contig of 715 bp in length
*   16028 16127: gap of      100 bp
*   16128    16828: contig of 701 bp in length
*   16829 16928: gap of      100 bp
*   16929    17634: contig of 706 bp in length
*   17635 17734: gap of      100 bp
*   17735    18453: contig of 719 bp in length
*   18454 18553: gap of      100 bp
*   18554    19272: contig of 719 bp in length
*   19273 19372: gap of      100 bp
*   19373    20047: contig of 675 bp in length
*   20048 20147: gap of      100 bp
*   20148    20861: contig of 714 bp in length
*   20862 20961: gap of      100 bp
*   20962    21670: contig of 709 bp in length
*   21671 21770: gap of      100 bp
*   21771    22479: contig of 709 bp in length
*   22480 22579: gap of      100 bp
*   22580    23299: contig of 720 bp in length
*   23300 23399: gap of      100 bp
*   23400    24127: contig of 728 bp in length
*   24128 24227: gap of      100 bp
*   24228    24937: contig of 710 bp in length
*   24938 25037: gap of      100 bp
*   25038    25750: contig of 713 bp in length
*   25751 25850: gap of      100 bp
*   25851    26524: contig of 674 bp in length
*   26525 26624: gap of      100 bp
*   26625    27338: contig of 714 bp in length
*   27339 27438: gap of      100 bp
*   27439    28163: contig of 725 bp in length
*   28164 28263: gap of      100 bp
*   28264    28988: contig of 725 bp in length
*   28989 29088: gap of      100 bp
*   29089    29801: contig of 713 bp in length
*   29802 29901: gap of      100 bp
*   29902    30602: contig of 701 bp in length
*   30603 30702: gap of      100 bp
*   30703    31412: contig of 710 bp in length
*   31413 31512: gap of      100 bp
*   31513    32238: contig of 726 bp in length
*   32239 32338: gap of      100 bp
*   32339    33068: contig of 730 bp in length
*   33069 33168: gap of      100 bp
*   33169    33887: contig of 719 bp in length
*   33888 33987: gap of      100 bp
*   33988    34697: contig of 710 bp in length
*   34698 34797: gap of      100 bp
*   34798    35508: contig of 711 bp in length
*   35509 35608: gap of      100 bp
*   35609    36320: contig of 712 bp in length
*   36321 36420: gap of      100 bp
*   36421    37113: contig of 693 bp in length
*   37114 37213: gap of      100 bp
*   37214    37934: contig of 721 bp in length
*   37935 38034: gap of      100 bp
*   38035    38725: contig of 691 bp in length
*   38726 38825: gap of      100 bp
*   38826    39537: contig of 712 bp in length
*   39538 39637: gap of      100 bp
*   39638    40342: contig of 705 bp in length
*   40343 40442: gap of      100 bp
*   40443    41151: contig of 709 bp in length
*   41152 41251: gap of      100 bp
*   41252    41966: contig of 715 bp in length
*   41967 42066: gap of      100 bp
*   42067    42774: contig of 708 bp in length
*   42775 42874: gap of      100 bp
*   42875    43580: contig of 706 bp in length
*   43581 43680: gap of      100 bp
*   43681    44360: contig of 680 bp in length
*   44361 44460: gap of      100 bp
*   44461    45162: contig of 702 bp in length
*   45163 45262: gap of      100 bp
*   45263    45957: contig of 695 bp in length
*   45958 46057: gap of      100 bp
*   46058    46774: contig of 717 bp in length
*   46775 46874: gap of      100 bp
*   46875    47570: contig of 696 bp in length
*   47571 47670: gap of      100 bp
*   47671    48380: contig of 710 bp in length
*   48381 48480: gap of      100 bp
*   48481    49195: contig of 715 bp in length
*   49196 49295: gap of      100 bp
*   49296    49998: contig of 703 bp in length
*   49999 50098: gap of      100 bp
*   50099    50820: contig of 722 bp in length
*   50821 50920: gap of      100 bp
*   50921    51624: contig of 704 bp in length
*   51625 51724: gap of      100 bp
*   51725    52433: contig of 709 bp in length
*   52434 52533: gap of      100 bp
```

```
            *    52534    53228: contig of 695 bp in length
            *    53229 53328: gap of      100 bp
            *    53329    54016: contig of 688 bp in length
            *    54017 54116: gap of      100 bp
            *    54117    54835: contig of 719 bp in length
            *    54836 54935: gap of      100 bp
            *    54936    55651: contig of 716 bp in length
            *    55652 55751: gap of      100 bp

  Query Match             9.1%;  Score 100;  DB 2;  Length 70244;
  Best Local Similarity   73.8%;  Pred. No. 1.2e-17;
  Matches  127;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     289 GCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCA 348
           |||||| || ||||| || || || |||||||||||| ||||| || |||||||| |||||
Db   20733 GCAAAAATATTTTGATTCTGGGGATTACAACATGGCTAAAGCAAAAATGAAGAACAAGCA 20674

Qy     349 GCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACA 408
            || || | |||||  || || ||||    ||| |||||||| ||||| ||||| || ||
Db   20673 ACTTCCTACTGCAGCTCCGGATAAGACGGAGGTCACTGGTGACCACATTCCCACTCCGCA 20614

Qy     409 GGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAA 460
           || || || ||  | ||| | || || ||  | |||||||| || || | |
Db   20613 AGACCTTCCTCAACGGAAGCCGTCCCTTGTTGCTAGCAAGCTGGCTGGCTGA 20562


Search completed: September 12, 2002, 15:34:27
Job time: 21131 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2002, 12:57:06 ; Search time 407.7 Seconds
(without alignments)
4644.976 Million cell updates/sec

Title: US-09-824-178-1
Perfect score: 1103
Sequence: 1 GTGACAGGAGCCGAAGCAGC.....AAAAAAAAAAGGGCGGCGCG 1103

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1103	100.0	1103	19	AAV45210	Human endosulfine
2	750.8	68.1	863	21	AAC98847	Human pancreatic c
3	456	41.3	553	24	ABK15891	Human lung tumour
4	456	41.3	808	22	AAH07746	Human cDNA clone (
5	456	41.3	1199	19	AAV45211	Human endosulfine
6	456	41.3	1208	22	AAH14942	Human cDNA sequenc
7	455.2	41.3	503	21	AAC01495	Human secreted pro
8	452	41.0	494	24	ABK15892	Human lung tumour
9	431.4	39.1	669	20	AAZ24516	Human lung tumor a
10	431.4	39.1	669	21	AAC65755	Human lung cancer-
11	271.4	24.6	454	19	AAV30285	Human phosphoprote
12	238	21.6	252	21	AAC17325	Human secreted pro
13	169.4	15.4	2498	22	AAI93443	Human polynucleoti
c 14	165.4	15.0	2544	22	AAK90275	Human digestive sy
c 15	165.4	15.0	2544	22	AAI57655	Human colorectal c
16	156.2	14.2	3316	19	AAV38840	Homo sapiens C4-2
17	121.8	11.0	341	21	AAC00778	Human secreted pro
18	121.8	11.0	350	21	AAC00776	Human secreted pro
19	114	10.3	295	21	AAC00775	Human secreted pro
20	111.8	10.1	407	21	AAC00777	Human secreted pro
21	67.8	6.1	91	21	AAC15418	Human secreted pro
22	65.4	5.9	343	14	AAQ60963	Human brain Expres
23	42.2	3.8	646	22	AAS29995	Human lung antigen
24	42.2	3.8	646	22	AAS28504	Genomic sequence #
25	41.8	3.8	929	21	AAC10860	Human secreted pro
26	41.8	3.8	3399	17	AAT05868	Chicken leucocytoz
27	41.4	3.8	12007	24	ABL32717	Human immune syste
28	40.6	3.7	882	23	AAS80806	DNA encoding novel
29	40.6	3.7	936	23	AAS71090	DNA encoding novel
30	40.4	3.7	714	22	AAH94306	Human foetal cDNA,
31	40.2	3.6	879	21	AAC01808	Human secreted pro
32	40.2	3.6	1380	23	ABL24141	Drosophila melanog
c 33	40.2	3.6	3651	23	ABL24140	Drosophila melanog
c 34	40.2	3.6	4794	23	ABL12632	Drosophila melanog
35	39.8	3.6	242	22	ABA48792	Human breast cell
36	39.8	3.6	242	22	ABA66709	Human foetal liver
37	39.8	3.6	242	22	ABA33768	Probe #12234 for g
38	39.8	3.6	242	22	AAK15132	Human brain expres
39	39.8	3.6	242	22	AAK40862	Human bone marrow
40	39.8	3.6	242	22	AAI21627	Probe #11560 for g
41	39.8	3.6	242	22	AAI46917	Probe #15603 used
42	39.8	3.6	242	22	AAI07322	Probe #7313 used t
43	39.8	3.6	467	22	ABA43686	Human breast cell
44	39.8	3.6	467	22	ABA54142	Human foetal liver
45	39.8	3.6	467	22	ABA23888	Probe #2354 for ge

ALIGNMENTS

RESULT 1
AAV45210
ID AAV45210 standard; cDNA; 1103 BP.
XX
AC AAV45210;
XX
DT 19-OCT-1998 (first entry)
XX
DE Human endosulfine A gene.
XX
KW ds; endosulfine; central nervous system; pancreatic beta-cell; pain;
KW insulin; cytoprotective agent; diabetes; epilepsy; depression; ischemia.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 107..460
FT /*tag= a
FT /product= "Endosulfine A protein"
XX
PN WO9830692-A2.
XX
PD 16-JUL-1998.
XX
PF 07-JAN-1998; 98WO-US00137.
XX
PR 07-JAN-1997; 97US-0779775.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Anderson KL, Roch J, Scott VE, Sullivan JP;
XX

```
DR   WPI; 1998-399140/34.
DR   P-PSDB; AAW61551.
XX
PT   Novel human endosulfine A and B - for screening for compounds
PT   modulating endosulfine receptor activity, useful as cytoprotective
PT   agent(s)
XX
PS   Claim 2; Page 57-58; 76pp; English.
XX
CC   The human endosulfine A and B genes encode proteins related to other
CC   members of the endosulfine family which have been shown to bind receptors
CC   from both the central nervous system and pancreatic beta-cells and to
CC   induce insulin secretion from a rodent beta-cell tumour line.  The
CC   protein can be used to screen for compounds that modulate endosulfine
CC   receptor activity, especially as a cytoprotective agent.
CC   Modulators of endosulfine activity may represent a treatment of
CC   diabetes, epilepsy, pain, depression and ischemia.
XX
SQ   Sequence 1103 BP; 313 A; 264 C; 267 G; 259 T; 0 other;


  Query Match             100.0%;  Score 1103;  DB 19;  Length 1103;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1103;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gtgacaggagccgaagcagcagcgcaggttgtccccgtttccccctcccccttcccttctc 60

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 cggttgccttcccgggcccccttacactccacagtcccggtcccgccatgtcccagaaaca 120

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 agaagaagagaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaaga 180

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aggtattctgcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctagg 240

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 acaaaagcctggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactt 300

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 tgactcaggagactacaacatggccaaagccaagatgaagaataagcagctgccaagtgc 360

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 aggaccagacaagaacctggtgactggtgatcacatccccaccccacaggatctgcccca 420

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 gagaaagtcctcgctcgtcaccagcaagcttgcggggtaacctgagcccccctctcctcc 480

Qy     481 CCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ccttcctcaaccactggacgtttatatattataggcagggatgaaatgggcacctagtca 540

Qy     541 GATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 gatcttctcagcttgctagccagaaatgactgtgattctgctgggggctgctgagaaggt 600

Qy     601 AATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 aatgtaggttgaaaaggggctctaagtttatttatttcgttagattgacacttccacaca 660

Qy     661 CTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ctccctgtagtccaggtagggcctagaaataggaaaggctaggattggataatgctgcaa 720

Qy     721 ATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 atgctttttttgtgtgagaaactggagagatgtgatttctcctttgggagagaatgtcc 780

Qy     781 CAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 caaaattgattaggctgagccttgggaatagtttggcaggtttaacatcccaaggctaac 840

Qy     841 CTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ctaacgtagttgggaaaggtagattgaatgagacatgttttctgtgcttctaagtgttct 900

Qy     901 GTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 gtcccttaggctgctattgcttcatgtttccattatggcaggtttagagaatccttaaaa 960

Qy     961 AGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 agaaaaattgacttgcttgcctaaaactacagtgccccccttagcctccattacttagtat 1020

Qy    1021 CTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACATTATTCACAAAA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 ctcttacagtttgctctggctctcaaataatataaagattgatgaacattattcacaaaa 1080

Qy    1081 AAAAAAAAAAAAAGGGCGGCGCG 1103
          |||||||||||||||||||||||
Db    1081 aaaaaaaaaaaaagggcggcgcg 1103


RESULT   2
AAC98847
ID   AAC98847 standard; cDNA; 863 BP.
XX
AC   AAC98847;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:75.
XX
KW   Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW   detection; diagnosis; identification; cytostatic; neuroprotective;
KW   nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW   antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW   linkage analysis; tissue identification; tissue typing; forensic;
KW   neural; immune system; muscular; reproductive; gastrointestinal;
KW   pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055320-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05989.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-579444/54.
DR   P-PSDB; AAB54082.
XX
PT   New nucleic acid that is a pancreatic cancer antigen for preventing,
PT   treating, or ameliorating a medical condition, particular pancreatic
PT   cancer, or for use in assays for diagnosing a pathological condition -
XX
PS   Claim 1; Page 549-550; 1379pp; English.
XX
CC   AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC   proteins, called pancreatic cancer antigens, given in AAB54008 to
CC   AAB54466. The human pancreatic cancer antigens have cytostatic,
```

bad date

```
CC   neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC   gynaecological, cardiant and antiinflammatory activities, and can be used
CC   in gene therapy. The polynucleotide and proteins can be used for
CC   preventing, treating, or ameliorating a medical condition or in assays
CC   for diagnosing a pathological condition or a susceptibility to one in a
CC   subject. Binding partners to the proteins and the activity of the
CC   proteins can be identified. The pancreatic cancer antigens can be used to
CC   detect, treat or prevent pancreatic disorders, especially cancer.
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 863 BP; 236 A; 215 C; 223 G; 186 T; 3 other;

  Query Match             68.1%;  Score 750.8;  DB 21;  Length 863;
  Best Local Similarity   96.7%;  Pred. No. 8.9e-204;
  Matches  829;  Conservative    0;  Mismatches   20;  Indels    8;  Gaps    6;

Qy       8 GAGCCGAAGCAGCAGCGCAGGTTGT-CCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTG 66
           | | |||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db       1 gnccgnaagcagcagcgcaggttgtaccccgtttcccctcccccttcccttctccggttg 60

Qy      67 CCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGA 126
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ccttcccgggccccttacactccacagtcccggtcccgccatgtcccagaaacaagaaga 120

Qy     127 AGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTAT 186
           ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db     121 agagaaccctgcggaggagaccggcgaggagaagcaggacacgca-gagaaagaaggtat 179

Qy     187 TCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAA 246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     180 tctgcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctaggacaaaa 239

Qy     247 GCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTC 306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     240 gcctggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactttgactc 299

Qy     307 AGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACC 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 aggagactacaacatggccaaagccaagatgaagaataagcagctgccaagtgcaggacc 359

Qy     367 AGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAA 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 agacaagaacctggtgactggtgatcacatccccaccccacaggatctgccccagagaaa 419

Qy     427 GTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCC 486
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 gtcctcgctcgtcaccagcaagcttgcggggtaacctgagccccctctcctccccttcc 479

Qy     487 TCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTT 546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 tcaaccactggacgtttatatattataggcagggatgaaatgggcacctagtcagatctt 539

Qy     547 CTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTA 606
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 ctcagcttgctagccagaaatgactgtgattctgctgggggctgctgagaaggtaatgta 599

Qy     607 GGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCT 666
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 ggttgaaaaggggctctaagtttatttatttcgttagattgacacttccacacactccct 659

Qy     667 GTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGCTT 726
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     660 gtagtccaggtagggcccagaaataggaaaggctaggattggataatgctgcaaatgctt 719

Qy     727 TTTTTGTGTGAGAAACT-GGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAAA 785
           ||||||||||||||||| |||||||||||||||||||||| |||||||||   || |||
Db     720 tttttgtgtgagaaactgggagagatgtgatttctccttttggggagagaattntcccaaa 779

Qy     786 TT--GATTAGGCTGAGCCTTGGGAATAGTTT--GGCAGGTTTAACATCCCAA-GGCTAAC 840
           ||    ||||| |||||||||||||||||||  |   | ||||||||||||| || ||||
Db     780 tttgatttagggtgagccttgggaatagttttggcaggttttaacatcccaagggttaac 839

Qy     841 CTAACGTAGTTGGGAAA 857
           ||||||||||| || ||
Db     840 ctaacgtagtttgggaa 856


RESULT   3
ABK15891
ID   ABK15891 standard; cDNA; 553 BP.
XX
AC   ABK15891;
XX
DT   14-MAR-2002  (first entry)
XX
DE   Human lung tumour protein encoding cDNA #3.
XX
KW   Human; lung tumour protein; lung cancer; immunostimulant; cytostatic; ss;
KW   PCR primer.
XX
OS   Homo sapiens.
XX
PN   WO200202623-A2.
XX
PD   10-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20975.
XX
PR   29-JUN-2000; 2000US-215696P.
PR   22-AUG-2000; 2000US-227142P.
PR   06-SEP-2000; 2000US-230481P.
PR   21-DEC-2000; 2000US-257729P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Mcneill PD,  Wantanabe Y,  Carter D,  Henderson RA,  Kalos MD;
XX
DR   WPI; 2002-106602/14.
XX
PT   Polynucleotides encoding lung tumour polypeptides, useful for treating
PT   lung cancer or stimulating an immune response -
XX
PS   Claim 1; Page 133; 277pp; English.
XX
CC   The invention relates to polynucleotides encoding lung tumour
CC   polypeptides. Compositions comprising the lung tumour polypeptides,
CC   polynucleotides, fusion proteins, antibodies to the polypeptides, T cell
CC   populations, or antigen presenting cells that express the lung tumour
CC   polypeptides, are useful for treating lung cancer or stimulating an
CC   immune response. The polynucleotides and polypeptides are also useful in
CC   the diagnosis and monitoring of lung cancer. Sequences ABK15889-AAS16343
CC   represent cDNA molecules encoding the polypeptides of the invention and
CC   PCR primers used to amplify the cDNA.
XX
SQ   Sequence 553 BP; 151 A; 164 C; 146 G; 92 T; 0 other;

  Query Match             41.3%;  Score 456;  DB 24;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 8.3e-120;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43 gtgacaggagccgaagcagcagcgcaggttgtccccgtttcccctcccccttcccttctc 102
```

```
Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103 cggttgccttcccgggcccttacactccacagtcccggtcccgccatgtcccagaaaca 162

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     163 agaagaagagaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaaga 222

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     223 aggtattctgcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctagg 282

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     283 acaaaagcctggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactt 342

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     343 tgactcaggagactacaacatggccaaagccaagatgaagaataagcagctgccaagtgc 402

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     403 aggaccagacaagaacctggtgactggtgatcacatccccaccccacaggatctgcccca 462

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
           |||||||||||||||||||||||||||||||||||
Db     463 gagaaagtcctcgctcgtcaccagcaagcttgcggg 498
```

RESULT 4
AAH07746
ID AAH07746 standard; cDNA; 808 BP.
XX
AC AAH07746;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:4581.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 4581; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 808 BP; 198 A; 218 C; 224 G; 159 T; 9 other;

Query Match 41.3%; Score 456; DB 22; Length 808;
Best Local Similarity 100.0%; Pred. No. 1e-119;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 gtgacaggagccgaagcagcagcgcaggttgtccccgtttccccctcccccttcccttctc 80

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 cggttgccttcccgggcccttacactccacagtcccggtcccgccatgtcccagaaaca 140

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 agaagaagagaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaaga 200

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 aggtattctgcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctagg 260

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 acaaaagcctggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactt 320

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 tgactcaggagactacaacatggccaaagccaagatgaagaataagcagctgccaagtgc 380

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 aggaccagacaagaacctggtgactggtgatcacatccccaccccacaggatctgcccca 440

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
           |||||||||||||||||||||||||||||||||||
Db     441 gagaaagtcctcgctcgtcaccagcaagcttgcggg 476
```

RESULT 5
AAV45211
ID AAV45211 standard; cDNA; 1199 BP.
XX
AC AAV45211;
XX
DT 19-OCT-1998 (first entry)
XX
DE Human endosulfine B gene.
XX
KW ds; endosulfine; central nervous system; pancreatic beta-cell; pain;
KW insulin; cytoprotective agent; diabetes; epilepsy; depression; ischemia.
XX
OS Homo sapiens.

```
XX
FH   Key             Location/Qualifiers
FT   CDS             107..472
FT                   /*tag= a
FT                   /product= "Endosulfine B protein"
XX
PN   WO9830692-A2.
XX
PD   16-JUL-1998.
XX
PF   07-JAN-1998;    98WO-US00137.
XX
PR   07-JAN-1997;    97US-0779775.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Anderson KL, Roch J, Scott VE, Sullivan JP;
XX
DR   WPI; 1998-399140/34.
DR   P-PSDB; AAW61552.
XX
PT   Novel human endosulfine A and B - for screening for compounds
PT   modulating endosulfine receptor activity, useful as cytoprotective
PT   agent(s)
XX
PS   Claim 2; Page 58; 76pp; English.
XX
CC   The human endosulfine A and B genes encode proteins related to other
CC   members of the endosulfine family which have been shown to bind receptors
CC   from both the central nervous system and pancreatic beta-cells and to
CC   induce insulin secretion from a rodent beta-cell tumour line.  The
CC   protein can be used to screen for compounds that modulate endosulfine
CC   receptor activity, especially as a cytoprotective agent.
CC   Modulators of endosulfine activity may represent a treatment of
CC   diabetes, epilepsy, pain, depression and ischemia.
XX
SQ   Sequence 1199 BP; 335 A; 288 C; 333 G; 243 T; 0 other;
```

```
  Query Match             41.3%;  Score 456;  DB 19;  Length 1199;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-119;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gtgacaggagccgaagcagcagcgcaggttgtccccgtttccccctcccccttcccttctc 60

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 cggttgccttcccgggccccttacactccacagtcccggtcccgccatgtcccagaaaca 120

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 agaagaagagaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaaga 180

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aggtattctgcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctagg 240

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 acaaaagcctggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactt 300

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 tgactcaggagactacaacatggccaaagccaagatgaagaataagcagctgccaagtgc 360

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 aggaccagacaagaacctggtgactggtgatcacatccccaccccacaggatctgcccca 420

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
          |||||||||||||||||||||||||||||||||||||
Db     421 gagaaagtcctcgctcgtcaccagcaagcttgcggg 456
```

```
RESULT   6
AAH14942
ID   AAH14942 standard; cDNA; 1208 BP.
XX
AC   AAH14942;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:12845.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 12845; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 1208 BP; 326 A; 295 C; 337 G; 250 T; 0 other;
```

```
  Query Match             41.3%;  Score 456;  DB 22;  Length 1208;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-119;
  Matches  456;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 gtgacaggagccgaagcagcagcgcaggttgtccccgtttccctcccccttcccttctc 80

Qy     61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 cggttgccttcccgggcccccttacactccacagtcccggtcccgccatgtcccagaaaca 140

Qy    121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 agaagaagagaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaaga 200

Qy    181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 aggtattctgcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctagg 260

Qy    241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 acaaaagcctggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactt 320

Qy    301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 tgactcaggagactacaacatggccaaagccaagatgaagaataagcagctgccaagtgc 380

Qy    361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 aggaccagacaagaacctggtgactggtgatcacatccccaccccacaggatctgcccca 440

Qy    421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
          |||||||||||||||||||||||||||||||||||
Db    441 gagaaagtcctcgctcgtcaccagcaagcttgcggg 476
```

RESULT 7
AAC01495
ID AAC01495 standard; cDNA; 503 BP.
XX
AC AAC01495;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 1493.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG01489.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 1493; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 503 BP; 143 A; 147 C; 133 G; 78 T; 2 other;

```
  Query Match          41.3%;  Score 455.2;  DB 21;  Length 503;
  Best Local Similarity 99.6%;  Pred. No. 1.3e-119;
  Matches  454;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     31 gtgacaggagccgaagcagcagcgcaggttgtccccgtttccctcccccttcccttctc 90

Qy     61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     91 cggttgccttcccgggcccccttacactccacagtcccggtcccgccatgtcccagaaaca 150

Qy    121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    151 agaagaagagaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaaga 210

Qy    181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    211 aggtattctgcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctagg 270

Qy    241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    271 acaaaagcctggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactt 330

Qy    301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    331 tgactcaggagactacaacatggccaaagccaagatgaagaataagcagctgccaagtgc 390

Qy    361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||
Db    391 aggaccagacaagaacctggtgactggtgatcacatccccamcccacaggatctgcccca 450

Qy    421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
          ||||||||||||||||||:||||||||||||||||
Db    451 gagaaagtcctcgctcgtcamcagcaagcttgcggg 486
```

RESULT 8
ABK15892
ID ABK15892 standard; cDNA; 494 BP.
XX
AC ABK15892;
XX
DT 14-MAR-2002 (first entry)
XX
DE Human lung tumour protein encoding cDNA #4.
XX
KW Human; lung tumour protein; lung cancer; immunostimulant; cytostatic; ss;
KW PCR primer.
XX
OS Homo sapiens.
XX
PN WO200202623-A2.
XX
PD 10-JAN-2002.
XX
PF 28-JUN-2001; 2001WO-US20975.
XX
PR 29-JUN-2000; 2000US-215696P.

```
PR   22-AUG-2000; 2000US-227142P.
PR   06-SEP-2000; 2000US-230481P.
PR   21-DEC-2000; 2000US-257729P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Mcneill PD,  Wantanabe Y,  Carter D,  Henderson RA,  Kalos MD;
XX
DR   WPI; 2002-106602/14.
XX
PT   Polynucleotides encoding lung tumour polypeptides, useful for treating
PT   lung cancer or stimulating an immune response -
XX
PS   Claim 1; Page 133; 277pp; English.
XX
CC   The invention relates to polynucleotides encoding lung tumour
CC   polypeptides. Compositions comprising the lung tumour polypeptides,
CC   polynucleotides, fusion proteins, antibodies to the polypeptides, T cell
CC   populations, or antigen presenting cells that express the lung tumour
CC   polypeptides, are useful for treating lung cancer or stimulating an
CC   immune response. The polynucleotides and polypeptides are also useful in
CC   the diagnosis and monitoring of lung cancer. Sequences ABK15889-AAS16343
CC   represent cDNA molecules encoding the polypeptides of the invention and
CC   PCR primers used to amplify the cDNA.
XX
SQ   Sequence 494 BP; 142 A; 146 C; 127 G; 79 T; 0 other;


  Query Match             41.0%;  Score 452;  DB 24;  Length 494;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-118;
  Matches  452;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43 gtgacaggagccgaagcagcagcgcaggttgtccccgtttcccctcccccttcccttctc 102

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103 cggttgccttcccgggcccccttacactccacagtcccggtcccgccatgtcccagaaaca 162

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     163 agaagaagagaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaaga 222

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     223 aggtattctgcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctagg 282

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     283 acaaaagcctggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactt 342

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     343 tgactcaggagactacaacatggccaaagccaagatgaagaataagcagctgccaagtgc 402

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     403 aggaccagacaagaacctggtgactggtgatcacatccccaccccacaggatctgcccca 462

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTG 452
          ||||||||||||||||||||||||||||||||
Db     463 gagaaagtcctcgctcgtcaccagcaagcttg 494


RESULT   9
AAZ24516
ID   AAZ24516 standard; cDNA; 669 BP.
XX
AC   AAZ24516;
XX
DT   07-DEC-1999  (first entry)
XX
DE   Human lung tumor associated polynucleotide.
XX
KW   Human; lung tumor; lung cancer; T cell stimulation; ss.
XX
OS   Homo sapiens.
XX
PN   WO9947674-A2.
XX
PD   23-SEP-1999.
XX
PF   17-MAR-1999;   99WO-US05798.
XX
PR   18-MAR-1998;   98US-0040802.
PR   18-MAR-1998;   98US-0040984.
PR   27-JUL-1998;   98US-0123912.
PR   27-JUL-1998;   98US-0123933.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Reed SG,  Wang T;
XX
DR   WPI; 1999-571839/48.
XX
PT   New isolated lung tumor polynucleotides, used to develop products for
PT   the treatment, prevention and monitoring the progression of lung cancer
PT   -
XX
PS   Claim 1; Page 66; 148pp; English.
XX
CC   The invention provides isolated human lung tumor nucleic acids and
CC   polypeptides. The polypeptides can be used for the treatment of lung
CC   cancer. The polypeptides and polynucleotides can be used to stimulate T
CC   cells or antigen presenting cells for use in the treatment of lung
CC   cancer. The polypeptides and monoclonal antibodies specific for the
CC   polypeptides can also be used to inhibit the development of lung cancer.
CC   Agents which bind the polypeptides can be used for detecting lung cancer
CC   and for monitoring the progression of lung cancer.
XX
SQ   Sequence 669 BP; 158 A; 208 C; 173 G; 119 T; 11 other;


  Query Match             99.1%;  Score 431.4;  DB 20;  Length 669;
  Best Local Similarity   99.1%;  Pred. No. 9.5e-113;
  Matches  443;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy      10 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 gccgaagcagcagcgcaggttgtccccgtttcccctcccccttcccttctccggttgcct 61

Qy      70 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 tcccgggcccccttacactccacagtcccggtcccgccatgtcccagaaacaagaagaaga 121

Qy     130 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 gaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaagaaggtattct 181

Qy     190 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 gcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctaggacaaaagcc 241

Qy     250 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 309
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     242 tggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactttgactcngg 301

Qy     310 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 369
           |||||||||||||||||||| |||||||||||||||||||||||||||| |||||||
Db     302 agactacaacatggccaaagccaacatgaagaataagcagctgccaagtgcangaccaga 361

Qy     370 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 429
           |||||||||||||||||||||||||||||||||||| |||| |||||||||||||||
Db     362 caagaacctggtgactggtgatcacatccccaccccacaggatctg-cccagagaaagtc 420
```

```
Qy    430 CTCGCTCGTCACCAGCAAGCTTGCGGG 456
          ||||||||||||||||||||||||||||
Db    421 ctcgctcgtcaccagcaagcttgcggg 447


RESULT  10
AAC65755
ID   AAC65755 standard; cDNA; 669 BP.
XX
AC   AAC65755;
XX
DT   21-FEB-2001  (first entry)
XX
DE   Human lung cancer-associated cDNA LST-S2-72.
XX
KW   Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW   vaccine; detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200061612-A2.
XX
PD   19-OCT-2000.
XX
PF   03-APR-2000; 2000WO-US08896.
XX
PR   02-APR-1999;   99US-0285479.
PR   17-DEC-1999;   99US-0466396.
PR   30-DEC-1999;   99US-0476496.
PR   10-JAN-2000; 2000US-0480884.
PR   22-FEB-2000; 2000US-0510376.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Fan L;
XX
DR   WPI; 2000-628399/60.
XX
PT   Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT   protein is used for detecting and monitoring progression of lung cancer
PT   in a patient -
XX
PS   Claim 1a; Page 99; 261pp; English.
XX
CC   This invention describes a novel isolated polypeptide (I) which
CC   comprising an immunogenic portion of a lung tumor protein or variant (P2)
CC   which have cytostatic activity. The polypeptides and polynucleotides are
CC   used in compositions and vaccines to inhibit the development of cancer,
CC   especially lung cancer, in a patient. Methods described in the invention
CC   can be used to monitor the progression of a cancer by carrying out the
CC   detection at subsequent time points and comparing the results from the
CC   different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
CC   are treated with P2, polynucleotides encoding P2 or antigen presenting
CC   cells expressing P2 and then administered to the patient to inhibit
CC   development of cancer.
XX
SQ   Sequence 669 BP; 158 A; 208 C; 173 G; 119 T; 11 other;


  Query Match             39.1%;  Score 431.4;  DB 21;  Length 669;
  Best Local Similarity   99.1%;  Pred. No. 9.5e-113;
  Matches  443;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy     10 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 69
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 gccgaagcagcagcgcaggttgtccccgtttccccteccccttccttctccggttgcct 61

Qy     70 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 tcccgggcccccttacactccacagtcccggtcccgccatgtcccagaaacaagaagaaga 121

Qy    130 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    122 gaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaagaaggtattct 181

Qy    190 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 gcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctaggacaaaagcc 241

Qy    250 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 309
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    242 tggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactttgactcngg 301

Qy    310 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 369
          |||||||||||||||||||||| |||||||||||||||||||||||||||| |||||||
Db    302 agactacaacatggccaaagccaacatgaagaataagcagctgccaagtgcangaccaga 361

Qy    370 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 429
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    362 caagaacctggtgactggtgatcacatccccaccccacaggatctg-cccagagaaagtc 420

Qy    430 CTCGCTCGTCACCAGCAAGCTTGCGGG 456
          ||||||||||||||||||||||||||||
Db    421 ctcgctcgtcaccagcaagcttgcggg 447


RESULT  11
AAV30285
ID   AAV30285 standard; DNA; 454 BP.
XX
AC   AAV30285;
XX
DT   28-SEP-1998  (first entry)
XX
DE   Human phosphoprotein hPSHP DNA.
XX
KW   hPSHP; phosphoprotein; neurotransmitter; signal transduction;
KW   human; Parkinson's disease; Huntington's diseases; diagnosis;
KW   drug screening; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             17..373
FT                   /*tag=  a
XX
PN   WO9818922-A1.
XX
PD   07-MAY-1998.
XX
PF   14-OCT-1997;   97WO-US18489.
XX
PR   28-OCT-1996;   96US-0739484.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Hillman JL;
XX
DR   WPI; 1998-272217/24.
DR   P-PSDB; AAW60214.
XX
PT   Human phosphoprotein polypeptide, hPSHP - useful for treating
PT   degenerative neuronal diseases e.g. Parkinson's and Huntington's,
PT   diagnosing hPSHP related diseases and drug screening
XX
PS   Claim 5; Page 49; 64pp; English.
XX
CC   This polynucleotide codes for hPSHP (see AAW60214) a novel human
CC   phosphoprotein that is believed to function in the signal
CC   transduction pathway of neurotransmitters in brain tissue and in
CC   cell development processes in peripheral organs, based on its
CC   chemical and structural homology with known phosphoproteins
CC   GI 162690 and GI 741603.  It is a consensus sequence derived from
CC   overlapping and/or extended Incyte clones 258512, 483293, 483619,
CC   485751 and 486742 (from hNT2 teratocarcinoma cell line cDNA library
CC   HNT2RAT01), 841590 (PROSNOT05), 1613311 (COLNTUT06), 918496
```

```
CC   (BRSTNOT04), 1518226 (BLADTUT04) and 1509755 (LUNGNOT14).  Incyte
CC   clone 25812 was initially identified from a retinoic acid treated
CC   HNT2RAT01 library through a computer-generated search for amino acid
CC   sequence alignments.  The invention also provides genetically
CC   engineered expression vectors, and host cells comprising nucleic
CC   acid sequences encoding hPSHP used in a claimed method for
CC   producing hPSHP, as well as probes and primers, antibodies,
CC   agonists and antagonists of hPSHP.  hPSHP or its fragments are
CC   useful therapeutically to regenerate and enhance survival of nerve
CC   cells to halt degenerative processes in brain diseases such as
CC   Parkinson's and Huntington's disease.  Polynucleotides encoding
CC   hPSHP are also useful therapeutically, e.g. to create antisense
CC   molecules to block the biological activity of hPSHP to treat
CC   diseases associated with hPSHP expression.  They may be used
CC   diagnostically e.g. to detect and quantify gene expression in
CC   biopsied tissues in which hPSHP expression is implicated (e.g.
CC   associated with neurological diseases) and to monitor hPSHP levels
CC   during therapeutic intervention.  The probes can also be used for
CC   detection of polynucleotide sequences encoding hPSHP in biological
CC   samples by hybridisation assay (the sample may be subjected to PCR
CC   prior to analysis) (claimed).  The products may also be used in
CC   drug screening.
XX
SQ   Sequence 454 BP; 120 A; 119 C; 129 G; 85 T; 1 other;


  Query Match             24.6%;  Score 271.4;  DB 19;  Length 454;
  Best Local Similarity   99.0%;  Pred. No. 3.3e-67;
  Matches  294;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    2;

Qy     162 AGGACACGCAGGAG-AAAGAAGGTATTCTGCCTGAG-AGAGCTGAAGAGGCAAAGCTAAA 219
           ||||||||||||||| |||||||||||||||||||| |||||||||||||||||||||||
Db      61 aggacacgcaggagcaaagaaggtattctgcctgagaagagctgaagaggcaaagctaaa 120

Qy     220 GGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACT 279
           ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     121 ggccaaatacccaagcctaggacaaaagcctggaggctccgactttctcatgaagagact 180

Qy     280 CCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAA 339
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ccagaaagggcaaaagtactttgactcaggagactacaacatggccaaagccaagatgaa 240

Qy     340 GAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCC 399
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 gaataagcagctgccaagtgcaggaccagacaagaacctggtgactggtgatcacatccc 300

Qy     400 CACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 caccccacaggatctgccccagagaaagtcctcgctcgtcaccagcaagcttgcggg 357


RESULT  12
AAC17325
ID   AAC17325 standard; cDNA; 252 BP.
XX
AC   AAC17325;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 21400.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 21400; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 252 BP; 67 A; 42 C; 62 G; 81 T; 0 other;


  Query Match             21.6%;  Score 238;  DB 21;  Length 252;
  Best Local Similarity   99.6%;  Pred. No. 8.2e-58;
  Matches  249;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy     747 GAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGG 806
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gagatgtgatttctccttttgggagagaatgtcccaaaattgattaggctgagccttggg 60

Qy     807 AATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTT-GGGAAAGGTAGATT 865
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      61 aatagtttggcaggtttaacatcccaaggctaacctaacgtagttggggaaaggtagatt 120

Qy     866 GAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCAT 925
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 gaatgagacatgttttctgtgcttctaagtgttctgtcccttaggctgctattgcttcat 180

Qy     926 GTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAA 985
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 gtttccattatggcaggtttagagaatccttaaaaagaaaaattgacttgcttgcctaaa 240

Qy     986 ACTACAGTGC 995
           ||||||||||
Db     241 actacagtgc 250


RESULT  13
AAI93443
ID   AAI93443 standard; cDNA; 2498 BP.
XX
AC   AAI93443;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 13503.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
```

```
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO13512.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 13503; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2498 BP; 735 A; 454 C; 464 G; 845 T; 0 other;


  Query Match             15.4%;  Score 169.4;  DB 22;  Length 2498;
  Best Local Similarity   70.3%;  Pred. No. 9.7e-38;
  Matches  227;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy    138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
Db     56 ccgaggcagcctccgcggaggagcagaaggaaatggaagataaagtgactagtccagaga 115

Qy    198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
Db    116 aagcagaagaagcaaaattaaaagcaagatatcctcatctgggacaaaagcctggaggtt 175

Qy    258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
Db    176 cagatttcttaaggaaacggttgcagaaagggcaaaaatattttgattctggggattaca 235

Qy    318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACC 377
Db    236 acatggctaaagcaaaaatgaagaacaagcaacttcctactgcagctccggataagacgg 295

Qy    378 TGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCG 437
Db    296 aggtcactggtgaccacattcccactccgcaagaccttcctcaacggaagccgtcccttg 355

Qy    438 TCACCAGCAAGCTTGCGGGGTAA 460
Db    356 ttgctagcaagctggctggctga 378


RESULT  14
AAK90275/c
ID   AAK90275 standard; DNA; 2544 BP.
XX
AC   AAK90275;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 3851.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01324.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
```

PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
PS Disclosure; SEQ ID NO 3851; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.
XX
SQ Sequence 2544 BP; 698 A; 653 C; 506 G; 687 T; 0 other;

Query Match 15.0%; Score 165.4; DB 22; Length 2544;
Best Local Similarity 96.6%; Pred. No. 1.3e-36;
Matches 169; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2531 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 2472

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2471 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 2412

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAG 175
          ||||||||||||||||||||||||||||||||||||||||||  | |  | ||
Db    2411 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGTGAAGTGGAAG 2357
```

RESULT 15
AAI57655/c
ID AAI57655 standard; DNA; 2544 BP.
XX
AC AAI57655;
XX
DT 19-OCT-2001 (first entry)
XX
DE Human colorectal cancer antigen coding sequence SEQ ID NO: 192.
XX
KW Human; colorectal cancer; colorectal cancer antigen; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155350-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01350.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-457727/49.
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the colon and rectum including colorectal cancers
PT and also for testing and detection e.g. diagnosis -
XX
PS Disclosure; SEQ ID NO: 192; 522pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a

```
CC   number of colorectal cancer antigens. These are shown in
CC   AAI57547-AAI57619 and AAM38569-AAM38641. These can be used in the
CC   diagnosis, prevention and treatment of cancer of the colon and/or rectum.
CC   The present sequence is a colorectal cancer antigen genomic sequence.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2544 BP; 698 A; 653 C; 506 G; 687 T; 0 other;


  Query Match             15.0%;  Score 165.4;  DB 22;  Length 2544;
  Best Local Similarity   96.6%;  Pred. No. 1.3e-36;
  Matches  169;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2531 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 2472

Qy     61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2471 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 2412

Qy    121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAG 175
          |||||||||||||||||||||||||||||||||||||||||||  | |  | ||
Db   2411 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGTGAAGTGGAAG 2357


RESULT  16
AAV38840
ID   AAV38840 standard; DNA; 3316 BP.
XX
AC   AAV38840;
XX
DT   27-OCT-1998  (first entry)
XX
DE   Homo sapiens C4-2 gene.
XX
KW   C4-2; tumour suppressor gene; cancer; treatment; glioblastoma;
KW   cell proliferative disorders; brain tumour; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..291
FT                   /*tag= a
FT                   /product= C4-2 tumour suppressor
XX
PN   WO9820021-A1.
XX
PD   14-MAY-1998.
XX
PF   07-NOV-1997;   97WO-US20408.
XX
PR   08-NOV-1996;   96US-0744905.
XX
PA   (NWBI-) NORTHWEST BIOTHERAPEUTICS LLC.
XX
PI   Boynton AL,  Murphy GP,  Sehgal A;
XX
DR   WPI; 1998-286852/25.
DR   P-PSDB; AAW62655.
XX
PT   Tumour suppressor, C4-2 - useful for treatment of proliferative
PT   diseases e.g. cancer
XX
PS   Example; Fig 8; 112pp; English.
XX
CC   The sequence is that of the C4-2 tumour suppressor gene.
CC   The gene, its antisense sequences and its encoded protein
CC   can be used in the treatment of cell proliferative diseases,
CC   e.g. cancer, in particular glioblastomas and other malignant
CC   brain tumours.
XX
SQ   Sequence 3316 BP; 1010 A; 565 C; 581 G; 1160 T; 0 other;


  Query Match             14.2%;  Score 156.2;  DB 19;  Length 3316;
  Best Local Similarity   71.3%;  Pred. No. 6.5e-34;
  Matches  206;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy    172 GGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCC 231
          |||  |  |||  | |   || |||| ||| ||||| |||||  |||| || | ||| ||
Db      3 ggaagataaagtgactagtccagagaaagcagaagaagcaaaattaaaagcaagatatcc 62

Qy    232 AAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCA 291
              || |||||||||||||||||  |  ||| | | |||  |  | ||||||||||||
Db     63 tcatctgggacaaaagcctggaggttcaaatttcttaaggaaacggttgcagaaagggca 122

Qy    292 AAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCT 351
          ||| || ||||| || || || ||||||||||| ||||| || |||||||| ||||| ||
Db    123 aaaatattttgattctggggattacaacatggctaaagcaaaaatgaagaacaagcaact 182

Qy    352 GCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGA 411
          || | |||||  || || ||||    ||| |||||||| ||||| ||||| || || ||
Db    183 tcctactgcagctccggataagacggaggtcactggtgaccacattcccactccgcaaga 242

Qy    412 TCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAA 460
          || || ||    ||| | || || ||  | |||||||| || || | |
Db    243 ccttcctcaactgaagccgtcccttgttgctagcaagctggctggctga 291


RESULT  17
AAC00778
ID   AAC00778 standard; cDNA; 341 BP.
XX
AC   AAC00778;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 776.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG00772.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 776; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
```

```
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 341 BP; 102 A; 76 C; 99 G; 61 T; 3 other;

  Query Match           11.0%;  Score 121.8;  DB 21;  Length 341;
  Best Local Similarity  70.7%;  Pred. No. 1.3e-24;
  Matches  162;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
           | ||||   ||  || |||| ||||| |      ||| | ||| | |   || ||||
Db     112 ccgaggcagcctccgcggaggagcagaaggaaatggaagataaagtgactagtccagaga 171

Qy     198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
            ||| ||||| ||||| |||| || | ||| ||    || ||||||||||||||||| |
Db     172 aagcagaagaagcaaaattaaaagcaagatatcctcatctgggacaaaagcctggaggtt 231

Qy     258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
           | || ||| | | |||  |  | |||||||||||||| || ||||| || || || ||||
Db     232 cagatttcttaaggaaacggttgcagaaagggcaaaaatattttgattctggggattaca 291

Qy     318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACC 366
           ||||||| ||||| || |||||||| ||||| || || | |||||  ||
Db     292 acatggctaaagcaaaaatgaagaacaagcaacttcctactgcagctcc 340


RESULT  18
AAC00776
ID   AAC00776 standard; cDNA; 350 BP.
XX
AC   AAC00776;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 774.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG00770.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 774; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 350 BP; 124 A; 63 C; 106 G; 57 T; 0 other;

  Query Match           11.0%;  Score 121.8;  DB 21;  Length 350;
  Best Local Similarity  70.7%;  Pred. No. 1.3e-24;
  Matches  162;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     138 CGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGA 197
           | ||||   ||  || |||| ||||| |      |||  |  |||  | |   || ||||
Db     121 ccgaggcagcctccgcggaggagcagaaggaaatggaagataaagtgactagtccagaga 180

Qy     198 GAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCT 257
            ||| ||||| |||||  |||| || | ||| ||    || ||||||||||||||||| |
Db     181 aagcagaagaagcaaaattaaaagcaagatatcctcatctgggacaaaagcctggaggtt 240

Qy     258 CCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACA 317
           | || ||| | | |||  |  | |||||||||||||| || ||||| || || || ||||
Db     241 cagatttcttaaggaaacggttgcagaaagggcaaaaatattttgattctggggattaca 300

Qy     318 ACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACC 366
           ||||||| ||||| || |||||||| ||||| || || | |||||  ||
Db     301 acatggctaaagcaaaaatgaagaacaagcaacttcctactgcagctcc 349


RESULT  19
AAC00775
ID   AAC00775 standard; cDNA; 295 BP.
XX
AC   AAC00775;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 773.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG00769.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 773; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
```

```
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 295 BP; 111 A; 48 C; 85 G; 51 T; 0 other;


  Query Match             10.3%;  Score 114;  DB 21;  Length 295;
  Best Local Similarity   68.8%;  Pred. No. 2.1e-22;
  Matches  172;  Conservative    0;  Mismatches   75;  Indels    3;  Gaps    1;

Qy    120 AAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGG---ACACGCAGGAGA 176
          | ||| ||  |||   |  ||| || | ||    |||||||| ||   |      |||
Db     45 aggaaaaaaggaagtgtttggacgacagcgagcgggagaagcggggtcaagaaatggaag 104

Qy    177 AAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCC 236
          |  |||  | |   || |||| ||| ||||| |||||  |||| || | ||| ||    |
Db    105 ataaagtgactagtccagagaaagcagaagaagcaaaattaaaagcaagatatcctcatc 164

Qy    237 TAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGT 296
          | ||||||||||||||||| || || ||| | | |||  |  | ||||||||||||| |
Db    165 tgggacaaaagcctggaggttcagatttcttaaggaaacggttgcagaaagggcaaaaat 224

Qy    297 ACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAA 356
          | ||||| || || || |||||||||| ||||| || |||||||| ||||| || || |
Db    225 attttgattctggggattacaacatggctaaagcaaaaatgaagaacaagcaacttccta 284

Qy    357 GTGCAGGACC 366
           |||||  ||
Db    285 ctgcagctcc 294


RESULT  20
AAC00777
ID   AAC00777 standard; cDNA; 407 BP.
XX
AC   AAC00777;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 775.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG00771.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 775; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 407 BP; 140 A; 81 C; 116 G; 70 T; 0 other;


  Query Match             10.1%;  Score 111.8;  DB 21;  Length 407;
  Best Local Similarity   73.3%;  Pred. No. 1e-21;
  Matches  143;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy    172 GGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCC 231
          |||  |  |||  | |   || |||| ||| ||||| |||||  |||| || | ||| ||
Db    212 ggaagataaagtgactagtccagagaaagcagaagaagcaaaattaaaagcaagatatcc 271

Qy    232 AAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCA 291
             || |||||||||||||||| || || ||| | | |||  |  | ||||||||||
Db    272 tcatctgggacaaaagcctggaggttcagatttcttaaggaaacggttgcagaaagggca 331

Qy    292 AAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCT 351
          ||| || ||||| || || || |||||||||| ||||| || |||||||| ||||| ||
Db    332 aaaatattttgattctggggattacaacatggctaaagcaaaaatgaagaacaagcaact 391

Qy    352 GCCAAGTGCAGGACC 366
            || | |||||  ||
Db    392 tcctactgcagctcc 406


RESULT  21
AAC15418
ID   AAC15418 standard; cDNA; 91 BP.
XX
AC   AAC15418;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 19493.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
```

```
XX
PS   Claim 1; SEQ ID 19493; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 91 BP; 13 A; 33 C; 20 G; 14 T; 11 other;


  Query Match             6.1%;  Score 67.8;  DB 21;  Length 91;
  Best Local Similarity   90.1%;  Pred. No. 1.7e-09;
  Matches   64;  Conservative    7;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           |:|||||||||||||||||||||||::|||||||||||||::||||||||||:|||
Db      21 gkgacaggagccgaagcagcagcgcagghwgtccccgtttccccwmccccttccctwctc 80

Qy      61 CGGTTGCCTTC 71
           |||:|||||||
Db      81 cggwtgccttc 91


RESULT  22
AAQ60963
ID   AAQ60963 standard; DNA; 343 BP.
XX
AC   AAQ60963;
XX
DT   16-MAR-1994  (first entry)
XX
DE   Human brain Expressed Sequence Tag EST01041.
XX
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal;
KW   cAMP-regulated phosphoprotein; signal transduction; ss.
XX
OS   Homo sapiens.
XX
PN   WO9316178-A.
XX
PD   19-AUG-1993.
XX
PF   12-FEB-1993;   93WO-US01294.
XX
PR   12-FEB-1992;   92US-0837195.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Adams MD,  Moreno RF,  Venter CJ;
XX
DR   WPI; 1993-272882/34.
XX
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
XX
PS   Example 4; Page 418; 500pp; English.
XX
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
CC   EST01041 was found to have 86.4 per cent identity with a
CC   cAMP-regulated phosphoprotein in the PIR database and has been
CC   classified in the "Signal Transduction" group of ESTs.
CC   See also Q59041-Q61440.
XX
SQ   Sequence 343 BP; 62 A; 104 C; 92 G; 84 T; 1 other;


  Query Match             5.9%;  Score 65.4;  DB 14;  Length 343;
  Best Local Similarity   98.5%;  Pred. No. 1.6e-08;
  Matches   66;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     390 ATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGC 449
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 atcacttccccaccccacaggatctgccccagagaaagtcctcgctcgtcaccagcaagc 60

Qy     450 TTGCGGG 456
           |||||||
Db      61 ttgcggg 67


RESULT  23
AAS29995
ID   AAS29995 standard; DNA; 646 BP.
XX
AC   AAS29995;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human lung antigen genomic DNA #65.
XX
KW   Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;
KW   chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW   cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW   ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;
KW   hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW   cerebrovascular disorder; nervous system disorder; bacterial infection;
KW   fungal infection; viral infection; ocular disorder; endocrine disorder;
KW   gastrointestinal disorder; renal disorder; respiratory disorder;
KW   wound healing; skin aging; organ transplantation; food preservative;
KW   tissue regeneration; anti-infertility; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200155303-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01301.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
```

PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-457723/49.
XX
PT Isolated polypeptide for treating, preventing and/or prognosing
PT respiratory disorders related to the lung including lung cancers and
PT also for testing and detection e.g. diagnosis -
XX
PS Claim 1; SEQ ID No 259; 507pp; English.
XX
CC Sequences AAS29931-AAS30164 represent genomic DNA molecules, which encode
CC the lung antigen polypeptides of the invention. Lung antigen polypeptides
CC and their associated polynucleotides are useful in the diagnosis,
CC treatment and prevention of various types of disorders in e.g. humans,
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
CC pathological condition can be determined by detecting the presence or
CC absence of a mutation in a lung antigen polynucleotide. The treatable
CC disorders include autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischaemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and

```
CC   respiratory disorders such as asthma and pleurisy. The polypeptides can
CC   also be used to aid wound healing, to prevent skin aging due to sunburn,
CC   to maintain organs before transplantation, to regenerate tissues and in
CC   chemotaxis. The polypeptides can also be used as a food additive or
CC   preservative to increase or decrease storage capabilities.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.

  Query Match             3.8%;  Score 42.2;  DB 22;  Length 646;
  Best Local Similarity   55.0%;  Pred. No. 0.091;
  Matches   83;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy      75 GGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACC 134
           | |||    |  ||  ||   |||   || ||||| || ||| | | |  ||||   |
Db      83 gtcccgccgccgtcgccatcgccgcggccaccatgcccaagagaaaggctgaaggagatg 142

Qy     135 CTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTG 194
           ||   |  || |  | | ||| |||| | ||  |||||   | |   | ||   || |||
Db     143 ctaaaggagataaagccaaggtgaaggacgaaccgcagagaagatccgctaggttgtctg 202

Qy     195 AGAGAGCTGAAGAGGCAAAGCTAAAGGCCAA 225
             | | |||      |||||| | || ||||
Db     203 ctaaacctgctcctccaaagccagagcccaa 233


RESULT  24
AAS28504
ID   AAS28504 standard; DNA; 646 BP.
XX
AC   AAS28504;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Genomic sequence #344 encoding for novel human respiratory antigen.
XX
KW   Human; respiratory antigen; respiratory disorder; throat disorder;
KW   lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;
KW   anti allergic; anti asthmatic; anti inflammatory; olfactory;
KW   respiratory active; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155448-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01333.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
```

```
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-476224/51.
XX
PT   Isolated polypeptide for treating, preventing and/ or prognosing
PT   disorders related to the respiratory system including respiratory
PT   cancers and also for testing and detection e.g. diagnosis -
XX
PS   Disclosure; SED ID No 938; 546pp; English.
XX
CC   The present invention relates to the isolation of novel human
CC   respiratory antigens (AAU17685-AAU17975), and cDNA and genomic
CC   sequences encoding for these polypeptides. The sequences of the
CC   invention are useful for preventing, treating and/or prognosing
CC   disorders related to the respiratory system including throat
CC   disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis),
CC   lung disorders e.g. pneumonia, allergic disorders e.g. asthma,
CC   pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of
CC   the respiratory tissues e.g. lung cancer. The polynucleotide sequences
CC   of the invention are useful in gene therapy and antisense therapy.
CC   AAS28161-AAS28764 represent genomic sequences encoding for novel
CC   human respiratory antigens.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 646 BP; 165 A; 170 C; 182 G; 129 T; 0 other;

  Query Match             3.8%;  Score 42.2;  DB 22;  Length 646;
  Best Local Similarity   55.0%;  Pred. No. 0.091;
  Matches   83;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy      75 GGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACC 134
           | |||    |  ||  ||   |||   || ||||| || ||| | | |  ||||   |
Db      83 gtcccgccgccgtcgccatcgccgcggccaccatgcccaagagaaaggctgaaggagatg 142

Qy     135 CTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTG 194
           ||   |  || |  | | ||| |||| | ||  |||||   | |   | ||   || |||
Db     143 ctaaaggagataaagccaaggtgaaggacgaaccgcagagaagatccgctaggttgtctg 202

Qy     195 AGAGAGCTGAAGAGGCAAAGCTAAAGGCCAA 225
             | | |||      |||||| | || ||||
Db     203 ctaaacctgctcctccaaagccagagcccaa 233


RESULT  25
AAC10860
ID   AAC10860 standard; cDNA; 929 BP.
XX
AC   AAC10860;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 14935.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 14935; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 929 BP; 286 A; 245 C; 273 G; 121 T; 4 other;

  Query Match             3.8%;  Score 41.8;  DB 21;  Length 929;
  Best Local Similarity   44.8%;  Pred. No. 0.14;
  Matches  151;  Conservative    2;  Mismatches  184;  Indels    0;  Gaps    0;

Qy     113 CAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAG 172
```

```
Db    256 cagtacgaagaagaagaggaagcccaggaggaagctgcagaagtccatgaggaagttcat 315

Qy    173 GAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCA 232

Db    316 gaaccagaggaagttcaagaagaggagaaaccgagasccaaactcactgctcctaagatc 375

Qy    233 AGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAA 292

Db    376 ccagaaggggagaaagtggacttcgatgacatccagaagaagcgtcagaacaaagaccta 435

Qy    293 AAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTG 352

Db    436 atggagctccaggccctcatcgacagccactttgaagcccggaagaaggaggaggaggag 495

Qy    353 CCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGAT 412

Db    496 ctggtcgctctcaaagagagaatcgagaagcgccgtgcagagagwgccwgcagccggtgc 555

Qy    413 CTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGC 449

Db    556 tgtccacagggagctccagcccttctcacactcgacc 592


RESULT  26
AAT05868
ID   AAT05868 standard; DNA; 3399 BP.
XX
AC   AAT05868;
XX
DT   14-AUG-1996  (first entry)
XX
DE   Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.
XX
KW   Chicken leucocytozoan; immunogen; recombinant vaccine; protection;
KW   immunisation; vaccination; ss.
XX
OS   Chicken leucocytozoan.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..3399
FT                   /*tag=  a
FT   misc_feature    1150..3218
FT                   /*tag=  b
FT                   /note= "fragment referred to in the claims, for
FT                          use as insert in a recombinant vaccine
FT                          against chicken leucocytozoan disease"
XX
PN   JP07284392-A.
XX
PD   31-OCT-1995.
XX
PF   19-APR-1994;   94JP-0080643.
XX
PR   19-APR-1994;   94JP-0080643.
XX
PA   (DOBU-) DOBUTSUYO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
PA   (KITA ) KITASATO KENKYUSHO SH.
XX
DR   WPI; 1996-006311/01.
DR   P-PSDB; AAR97866.
XX
PT   Chicken leucocytozoan immunogenic protein  - used in a recombinant
PT   vaccine against chicken leucocytozoan disease
XX
PS   Claim 6; Page 6-9; 35pp; Japanese.
XX
CC   AAT05868 encodes a chicken leucocytozoan immunogenic protein, this DNA
CC   or a fragment of it can be used in a recombinant vaccine to immunise
CC   against chicken leucocytozoan disease. The DNA is used in a vector
CC   and operatively linked to an expression regulatory sequence as in
CC   standard practice.
XX

SQ   Sequence 3399 BP; 1577 A; 508 C; 798 G; 516 T; 0 other;


  Query Match             3.8%;  Score 41.8;  DB 17;  Length 3399;
  Best Local Similarity   49.1%;  Pred. No. 0.28;
  Matches  139;  Conservative    0;  Mismatches  142;  Indels    2;  Gaps    1;

Qy    114 AGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGG 173

Db   2753 aggaagaagaagaagaagaagatgaggaagaggaagaagaagaagaagaagatgaggaag 2812

Qy    174 AGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAA 233

Db   2813 aggaagaagaagaagaaaatgaggaagaagaagaagaagaaaataaggaagaagaagaag 2872

Qy    234 GCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAA 293

Db   2873 aagaaaaagaagagcatgaagaagaag--taacacatgaagaagaagaagaaaaagtaac 2930

Qy    294 AGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGC 353

Db   2931 acatgaagaagaagaaaaagtaacacatgaagaagaagaaaaatgtaacatatgaagaaga 2990

Qy    354 CAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACAT 396

Db   2991 agaagaaaaagtaacacatgaagaagaagaaaaagtaacacat 3033


RESULT  27
ABL32717
ID   ABL32717 standard; DNA; 12007 BP.
XX
AC   ABL32717;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 690.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 690; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
```

```
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 12007 BP; 3266 A; 114 C; 2487 G; 6140 T; 0 other;


  Query Match             3.8%;  Score 41.4;  DB 24;  Length 12007;
  Best Local Similarity   56.1%;  Pred. No. 0.68;
  Matches   78;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy     854 GAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTG 913
           ||||||   | ||||  ||| |||||||| | |  || || || || | |   |||   |
Db    7107 gaaaggggaagtgaaaaagaaatgttttgttttattataggttttttttttttttattttt 7166

Qy     914 CTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACT 973
           | ||  ||  | |||   || | | || |||||||  |  |  || | |     ||
Db    7167 ttttttttagttttttttttttagaagatttagagtttttagaataaatttggtgtag 7226

Qy     974 TGCTTGCCTAAAACTACAG 992
           |   ||  ||||  || ||
Db    7227 tttatgtttaaatttatag 7245


RESULT  28
AAS80806
ID   AAS80806 standard; cDNA; 882 BP.
XX
AC   AAS80806;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #16610.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG16619.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 16610; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 882 BP; 322 A; 165 C; 245 G; 150 T; 0 other;


  Query Match             3.7%;  Score 40.6;  DB 23;  Length 882;
  Best Local Similarity   47.5%;  Pred. No. 0.31;
  Matches  121;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy     117 AACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGA 176
           || ||||||||||  |    | |||||||   |  |||||| || ||||   | | ||
Db     119 aagaagaagaagaagaggaggaggaggaggaggaggaggaggaggaggaggaagaag 178

Qy     177 AAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCC 236
           ||||||           ||   ||  ||||| | | |   | | |   || | |
Db     179 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 238

Qy     237 TAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGT 296
           || | || |    | ||    ||          | |   |   ||   ||    || ||
Db     239 aagaagaagaaggagaagaagaagaaggagaagaggaagaagatcaactagacacaatgt 298

Qy     297 ACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAA 356
             | ||||||  |    ||  ||    || |||    |  ||| | |||    ||   |
Db     299 tatgggactcatcaacaaacctcaccaacacagcgctcagcaaggagaagaccatgttca 358

Qy     357 GTGCAGGACCAGACA 371
           || || |  |  | |
Db     359 gttcaagggccaaga 373


RESULT  29
AAS71090
ID   AAS71090 standard; cDNA; 936 BP.
XX
AC   AAS71090;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #6894.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG06903.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
```

```
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 6894; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 936 BP; 335 A; 174 C; 263 G; 164 T; 0 other;

  Query Match             3.7%;  Score 40.6;  DB 23;  Length 936;
  Best Local Similarity   47.5%;  Pred. No. 0.32;
  Matches  121;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy     117 AACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGA 176
           || |||||||||| |    | |||||||   |  |||||| || ||||   | | ||
Db     173 aagaagaagaagaagaggaggaggaggaggaggaggaggaggaggaggaggaagaag 232

Qy     177 AAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCC 236
           ||||||          ||   ||  ||||| | | |   | | |   || |   |
Db     233 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 292

Qy     237 TAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGT 296
           || | || |    | ||     ||       | |   |   ||   ||    || ||
Db     293 aagaagaagaaggagaagaagaagaaggagaagaagatcaactagacacaatgt 352

Qy     297 ACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAA 356
             |  ||||||  |    ||  ||    || |||    |  ||| | |||    ||   |
Db     353 tatgggactcatcaacaaacctcaccaacacagcgctcagcaaggagaagaccatgttca 412

Qy     357 GTGCAGGACCAGACA 371
           || || |  |  | |
Db     413 gttcaagggccaaga 427


RESULT  30
AAH94306
ID   AAH94306 standard; cDNA; 714 BP.
XX
AC   AAH94306;
XX
DT   05-OCT-2001  (first entry)
XX
DE   Human foetal cDNA, SEQ ID NO: 835.
XX
KW   Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;
KW   nootropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
KW   gene therapy; antisense therapy; cancer; immune disorder;
KW   growth disorder; osteoporosis; thrombolytic disorder;
KW   nervous system disorder; inflammation; expressed sequence tag; EST; ss.
XX
OS   Homo sapiens.
XX
PN   WO200155339-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02723.
XX
PR   25-JAN-2000; 2000US-0491404.
PR   15-SEP-2000; 2000US-0663870.
PR   06-NOV-2000; 2000US-0707351.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Yeung G,  Ford JE,  Boyle BJ,  Arterburn MC,  Drmanac RA,  Tang YT;
PI   Liu C,  Asundi V,  Zhou P,  Werhman T;
XX
DR   WPI; 2001-465571/50.
DR   P-PSDB; AAM06631.
XX
PT   Novel fetal proteins useful for the treatment and diagnosis of diseases
PT   associated with dysfunction of the protein e.g. cancers, immune
PT   disorders, growth disorders, thrombolytic disorders, nervous system
PT   disorders and inflammation -
XX
PS   Claim 1; Page 497; 715pp; English.
XX
CC   The invention relates to novel foetal polypeptides encoded by
CC   polynucleotides comprising one of 477 sequences fully defined in the
CC   specification. The foetal polynucleotides and polypeptides are
CC   useful in the treatment and diagnosis of diseases such as cancers,
CC   immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
CC   disorders, nervous system disorders and inflammation. The present
CC   sequence was assembled using an expressed sequence tag (EST) found
CC   to be expressed in human foetal tissue cDNA libraries as the seed.
XX
SQ   Sequence 714 BP; 255 A; 119 C; 154 G; 186 T; 0 other;

  Query Match             3.7%;  Score 40.4;  DB 22;  Length 714;
  Best Local Similarity   56.9%;  Pred. No. 0.31;
  Matches   74;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy     117 AACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGA 176
           || ||| | |||||||      ||||||||  |  |||||||||| ||||   | |  |||
Db     186 aagaaggataagagaaggagaaggaggagaaggatgaggagaaggaggaggggggaataga 245

Qy     177 AAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCC 236
           |   ||   |   |   || ||||    |||   | ||  ||||   |||  |  |||
Db     246 agagagaagtggagagagacagagagagagaccgagagagaaagaaaaaaggcataaggg 305

Qy     237 TAGGACAAAA 246
           | |  ||||
Db     306 aaaggaaaaa 315


RESULT  31
AAC01808
ID   AAC01808 standard; cDNA; 879 BP.
XX
AC   AAC01808;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 1806.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
```

```
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG01802.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 1806; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 879 BP; 268 A; 236 C; 259 G; 113 T; 3 other;


  Query Match             3.6%;  Score 40.2;  DB 21;  Length 879;
  Best Local Similarity   47.4%;  Pred. No. 0.4;
  Matches  120;  Conservative    0;  Mismatches  133;  Indels    0;  Gaps    0;

Qy     101 CCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAG 160
           ||| || |  ||  |    | || ||||   | |  | |||||||     ||| ||  |
Db     554 cccaccttcaccatgtctgacgaggaagttgaacaggtggaggagcagtacgaagaagaa 613

Qy     161 CAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAG 220
           ||||  | ||||||| |||| |   |||     |||             | || || |
Db     614 gaggaagcccaggaggaagaggaagttcaagaagaggagaaaccgagacccaaactcact 673

Qy     221 GCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTC 280
           ||     |          |||  | ||    |    |   ||| |||  | |||| | |
Db     674 gctcctaagatcccagaaggggagaaagtggacttcgatgacatccagaagaagcgtcag 733

Qy     281 CAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAG 340
           | ||||  | || | |  |  |  |      | ||| |       |||||  || ||||
Db     734 aacaaagacctaatggagctccaggccctcatcgacagccactttgaagcccggaagaag 793

Qy     341 AATAAGCAGCTGC 353
           |  || ||  ||
Db     794 gaggaggaggagc 806


RESULT  32
ABL24141
ID   ABL24141 standard; DNA; 1380 BP.
XX
AC   ABL24141;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 23896.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 23896; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1380 BP; 308 A; 392 C; 412 G; 268 T; 0 other;


  Query Match             3.6%;  Score 40.2;  DB 23;  Length 1380;
  Best Local Similarity   60.6%;  Pred. No. 0.5;
  Matches   66;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     113 CAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAG 172
           |||   ||||  |  ||||| ||  ||  || | ||| || |  |   |||     || |
Db     175 cagccccaagtggcggagaagccgccgccggtggccgtcgcgctgctccagccgcagcgg 234

Qy     173 GAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGG 221
           |||||  | ||  ||| |||| ||||    ||| |||   ||| | ||||
Db     235 gagaacggagtgattgtgcccgagaagtatgacgagcagaagatcaagg 283


RESULT  33
ABL24140/c
ID   ABL24140 standard; DNA; 3651 BP.
XX
AC   ABL24140;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 23893.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
```

```
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 23893; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3651 BP; 892 A; 907 C; 851 G; 1001 T; 0 other;

  Query Match             3.6%;  Score 40.2;  DB 23;  Length 3651;
  Best Local Similarity   60.6%;  Pred. No. 0.82;
  Matches   66;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     113 CAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAG 172
           |||   ||||  |  ||||| ||  ||  || | ||| || |  |   |||     || |
Db    2477 CAGCCCCAAGTGGCGGAGAAGCCGCCGCCGGTGGCCGTCGCGCTGCTCCAGCCGCAGCGG 2418

Qy     173 GAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGG 221
           ||||| | ||  ||| |||| ||||    ||| |||   ||| | ||||
Db    2417 GAGAACGGAGTGATTGTGCCCGAGAAGTATGACGAGCAGAAGATCAAGG 2369


RESULT  34
ABL12632/c
ID   ABL12632 standard; cDNA; 4794 BP.
XX
AC   ABL12632;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 32378.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB68529.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 32378; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4794 BP; 1200 A; 1079 C; 1134 G; 1381 T; 0 other;

  Query Match             3.6%;  Score 40.2;  DB 23;  Length 4794;
  Best Local Similarity   60.6%;  Pred. No. 0.94;
  Matches   66;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     113 CAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAG 172
           |||   ||||  |  ||||| ||  ||  || | ||| || |  |   |||     || |
Db     445 CAGCCCCAAGTGGCGGAGAAGCCGCCGCCGGTGGCCGTCGCGCTGCTCCAGCCGCAGCGG 386

Qy     173 GAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGG 221
           ||||| | ||  ||| |||| ||||    ||| |||   ||| | ||||
Db     385 GAGAACGGAGTGATTGTGCCCGAGAAGTATGACGAGCAGAAGATCAAGG 337


RESULT  35
ABA48792
ID   ABA48792 standard; DNA; 242 BP.
XX
AC   ABA48792;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human breast cell single exon nucleic acid probe #7487.
XX
KW   Human; microarray; single exon probe; gene expression; breast;
KW   disease; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157271-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00662.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-496933/54.
XX
PT   New spatially-addressable set of single exon nucleic acid probes,
PT   useful for measuring gene expression in sample derived from human
PT   breast, comprises number of single exon nucleic acid probes  -
```

```
XX
PS   Claim 4; SEQ ID NO 7487; 327pp + sequence listing; English.
XX
CC   The invention relates to a spatially-addressable set of single exon
CC   nucleic acid probes for measuring gene expression in a sample derived
CC   from human breast and BT 474 cells. The method involves contacting
CC   the probes with a collection of detectably labelled nucleic acids
CC   derived from mRNA of human breast, and then measuring the label
CC   bound to each probe of the microarray. The probes are useful for
CC   verifying the expression of regions of genomic DNA predicted to
CC   encode proteins. They are useful for gene discovery, and for
CC   determining predisposition and/or prognosing breast disease. Gene
CC   expression analysis is useful for assessing the toxicity of chemical
CC   agents on cells. The microarray of this invention presents a far greater
CC   diversity of probes for measuring gene expression, with far less bias
CC   than expressed sequence tag microarrays. The method is suitable for
CC   rapid production of functional information from genomic sequence. The
CC   present sequence is a single exon nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 242 BP; 118 A; 16 C; 89 G; 19 T; 0 other;


  Query Match             3.6%;  Score 39.8;  DB 22;  Length 242;
  Best Local Similarity   74.6%;  Pred. No. 0.27;
  Matches   50;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy    114 AGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGG 173
          |||| ||||||||||||||    | ||||||    |  ||||||||| |||| | | |||
Db     36 agaatcaagaagaagagaagaaggaggaggaagaggaggaggagaaggaggagaagaagg 95

Qy    174 AGAAAGA 180
          ||||  ||
Db     96 agaagga 102


RESULT  36
ABA66709
ID   ABA66709 standard; DNA; 242 BP.
XX
AC   ABA66709;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #15014.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 4; SEQ ID NO 15014; 639pp + sequence listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
CC   measuring and displaying gene expression in samples derived from human
CC   fetal liver. The present sequence is a single exon nucleic acid
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 242 BP; 118 A; 16 C; 89 G; 19 T; 0 other;


  Query Match             3.6%;  Score 39.8;  DB 22;  Length 242;
  Best Local Similarity   74.6%;  Pred. No. 0.27;
  Matches   50;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy    114 AGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGG 173
          |||| ||||||||||||||    | ||||||    |  ||||||||| |||| | | |||
Db     36 agaatcaagaagaagagaagaaggaggaggaagaggaggaggagaaggaggagaagaagg 95

Qy    174 AGAAAGA 180
          ||||  ||
Db     96 agaagga 102


RESULT  37
ABA33768
ID   ABA33768 standard; DNA; 242 BP.
XX
AC   ABA33768;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Probe #12234 for gene expression analysis in human heart cell sample.
XX
KW   Human; gene expression; heart; microarray; vascular system; probe;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00666.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts -
XX
PS   Claim 4; SEQ ID No 12234; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart. The
```

CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 242 BP; 118 A; 16 C; 89 G; 19 T; 0 other;

Query Match 3.6%; Score 39.8; DB 22; Length 242;
Best Local Similarity 74.6%; Pred. No. 0.27;
Matches 50; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 114 AGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGG 173
|||| |||||||||||||| | ||||||| | ||||||||| |||| | | |||
Db 36 agaatcaagaagaagagaagaaggaggaggaagaggaggaggagaaggaggagaagaagg 95

Qy 174 AGAAAGA 180
|||| ||
Db 96 agaagga 102

RESULT 38
AAK15132
ID AAK15132 standard; DNA; 242 BP.
XX
AC AAK15132;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 15123.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 15123; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 242 BP; 118 A; 16 C; 89 G; 19 T; 0 other;

Query Match 3.6%; Score 39.8; DB 22; Length 242;
Best Local Similarity 74.6%; Pred. No. 0.27;
Matches 50; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 114 AGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGG 173
|||| |||||||||||||| | ||||||| | ||||||||| |||| | | |||
Db 36 agaatcaagaagaagagaagaaggaggaggaagaggaggaggagaaggaggagaagaagg 95

Qy 174 AGAAAGA 180
|||| ||
Db 96 agaagga 102

RESULT 39
AAK40862
ID AAK40862 standard; DNA; 242 BP.
XX
AC AAK40862;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 15419.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 15419; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 242 BP; 118 A; 16 C; 89 G; 19 T; 0 other;

Query Match 3.6%; Score 39.8; DB 22; Length 242;
Best Local Similarity 74.6%; Pred. No. 0.27;
Matches 50; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
Qy    114 AGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGG 173
          |||| ||||||||||||||    | ||||||    |  ||||||||| |||| | | |||
Db     36 agaatcaagaagaagagaagaaggaggaggaagaggaggaggagaaggaggagaagaagg 95

Qy    174 AGAAAGA 180
          |||| ||
Db     96 agaagga 102
```

RESULT 40
AAI21627
ID AAI21627 standard; DNA; 242 BP.
XX
AC AAI21627;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #11560 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 11560; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 242 BP; 118 A; 16 C; 89 G; 19 T; 0 other;

Query Match 3.6%; Score 39.8; DB 22; Length 242;
Best Local Similarity 74.6%; Pred. No. 0.27;
Matches 50; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
Qy    114 AGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGG 173
          |||| ||||||||||||||    | ||||||    |  ||||||||| |||| | | |||
Db     36 agaatcaagaagaagagaagaaggaggaggaagaggaggaggagaaggaggagaagaagg 95

Qy    174 AGAAAGA 180
          |||| ||
Db     96 agaagga 102
```

Search completed: September 12, 2002, 16:14:20
Job time: 11834 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2002, 10:42:36 ; Search time 104.12 Seconds
(without alignments)
2602.130 Million cell updates/sec

Title: US-09-824-178-1
Perfect score: 1103
Sequence: 1 GTGACAGGAGCCGAAGCAGC.....AAAAAAAAAAGGGCGGCGCG 1103

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	431.4	39.1	669	4	US-09-040-984-16	Sequence 16, Appl
2	431.4	39.1	669	4	US-09-123-912-16	Sequence 16, Appl
3	271.4	24.6	454	2	US-08-739-484-2	Sequence 2, Appli
c 4	64.4	5.8	7218	1	US-08-232-463-14	Sequence 14, Appl
c 5	38	3.4	770	4	US-09-385-982-542	Sequence 542, App
6	37.4	3.4	3238	4	US-08-123-934A-5	Sequence 5, Appli
7	37.4	3.4	3238	5	PCT-US94-10080-5	Sequence 5, Appli
8	37	3.4	289	4	US-09-007-005-17	Sequence 17, Appl
9	37	3.4	289	4	US-09-244-796-17	Sequence 17, Appl
c 10	37	3.4	16442	3	US-08-781-891-208	Sequence 208, App
11	35.6	3.2	407	1	US-08-132-649-9	Sequence 9, Appli
12	35.6	3.2	407	3	US-08-767-579-9	Sequence 9, Appli
13	35.6	3.2	1265	1	US-08-132-649-5	Sequence 5, Appli
14	35.6	3.2	1265	3	US-08-767-579-5	Sequence 5, Appli
c 15	35	3.2	5183	1	US-08-459-568-3	Sequence 3, Appli
c 16	35	3.2	5183	2	US-08-399-411-3	Sequence 3, Appli
c 17	35	3.2	5868	3	US-08-516-859A-3	Sequence 3, Appli
c 18	34.4	3.1	420	2	US-08-557-892-5	Sequence 5, Appli
c 19	34.4	3.1	420	2	US-08-387-858A-5	Sequence 5, Appli
c 20	34.4	3.1	420	4	US-09-294-384B-5	Sequence 5, Appli
c 21	34.2	3.1	2233	1	US-08-145-705A-1	Sequence 1, Appli
22	34	3.1	1289	4	US-09-156-809-3	Sequence 3, Appli
23	34	3.1	6755	3	US-08-931-999-4	Sequence 4, Appli
c 24	33.8	3.1	669	4	US-09-040-984-16	Sequence 16, Appl
c 25	33.8	3.1	669	4	US-09-123-912-16	Sequence 16, Appl
26	33.4	3.0	1968	2	US-08-937-540-3	Sequence 3, Appli
c 27	33.4	3.0	51259	3	US-08-781-891-209	Sequence 209, App
c 28	33.2	3.0	32207	2	US-08-770-379-20	Sequence 20, Appl
c 29	33.2	3.0	32207	4	US-08-757-669A-20	Sequence 20, Appl
c 30	33.2	3.0	32207	4	US-09-230-371A-20	Sequence 20, Appl
31	33	3.0	1284	4	US-09-286-529-6	Sequence 6, Appli
32	33	3.0	1804	2	US-08-504-459-5	Sequence 5, Appli
33	33	3.0	2277	1	US-08-676-967-2	Sequence 2, Appli
34	33	3.0	2277	1	US-08-676-974-2	Sequence 2, Appli
35	33	3.0	2277	2	US-09-098-487-2	Sequence 2, Appli
36	32.8	3.0	547	1	US-08-131-365B-41	Sequence 41, Appl
37	32.8	3.0	547	2	US-08-668-123-41	Sequence 41, Appl
38	32.8	3.0	1734	4	US-08-630-915A-23	Sequence 23, Appl
39	32.8	3.0	2301	1	US-08-306-691B-23	Sequence 23, Appl
40	32.8	3.0	2301	5	PCT-US93-06251-78	Sequence 78, Appl
41	32.8	3.0	2470	4	US-09-091-725-18	Sequence 18, Appl
42	32.6	3.0	519	1	US-08-318-905-18	Sequence 18, Appl
43	32.6	3.0	519	1	US-08-483-232-18	Sequence 18, Appl
44	32.6	3.0	519	1	US-08-483-140-18	Sequence 18, Appl
45	32.6	3.0	519	2	US-08-485-938A-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-09-040-984-16
; Sequence 16, Application US/09040984
; Patent No. 6210883
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Wang, TongTong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS
; TITLE OF INVENTION: OF LUNG CANCER
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,984
; FILING DATE: 18-MAR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.456
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-282-6031
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-040-984-16

Query Match 39.1%; Score 431.4; DB 4; Length 669;
Best Local Similarity 99.1%; Pred. No. 6.5e-127;
Matches 443; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 10 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 69
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 61

```
Qy      70 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 121

Qy     130 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 181

Qy     190 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 241

Qy     250 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     242 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCNGG 301

Qy     310 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 369
           |||||||||||||||||||||| |||||||||||||||||||||||||| |||||||
Db     302 AGACTACAACATGGCCAAAGCCAACATGAAGAATAAGCAGCTGCCAAGTGCANGACCAGA 361

Qy     370 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 429
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     362 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTG-CCCAGAGAAAGTC 420

Qy     430 CTCGCTCGTCACCAGCAAGCTTGCGGG 456
           ||||||||||||||||||||||||||
Db     421 CTCGCTCGTCACCAGCAAGCTTGCGGG 447
```

```
RESULT   2
US-09-123-912-16
; Sequence 16, Application US/09123912A
; Patent No. 6312695
; GENERAL INFORMATION:
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Wang, Tongtong
;  TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
;  FILE REFERENCE:  210121.455C1
;  CURRENT APPLICATION NUMBER: US/09/123,912A
;  CURRENT FILING DATE:  1998-07-27
;  PRIOR APPLICATION NUMBER: 09/040,802
;  PRIOR FILING DATE:  1998-03-18
;  NUMBER OF SEQ ID NOS:  114
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 16
;   LENGTH: 669
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (299)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (354)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (483)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (555)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (571)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (573)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (577)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (642)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (651)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (662)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (667)
;   OTHER INFORMATION: Where n is a, c, g or t
US-09-123-912-16


  Query Match             39.1%;  Score 431.4;  DB 4;  Length 669;
  Best Local Similarity   99.1%;  Pred. No. 6.5e-127;
  Matches  443;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy      10 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 gccgaagcagcagcgcaggttgtccccgtttcccctcccccttccctttctccggttgcct 61

Qy      70 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 tcccgggcccctacactccacagtcccggtcccgccatgtcccagaaacaagaagaaga 121

Qy     130 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 gaaccctgcggaggagaccggcgaggagaagcaggacacgcaggagaaagaaggtattct 181

Qy     190 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 gcctgagagagctgaagaggcaaagctaaaggccaaatacccaagcctaggacaaaagcc 241

Qy     250 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     242 tggaggctccgacttcctcatgaagagactccagaaagggcaaaagtactttgactcngg 301

Qy     310 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 369
           |||||||||||||||||||||| |||||||||||||||||||||||||| |||||||
Db     302 agactacaacatggccaaagccaacatgaagaataagcagctgccaagtgcangaccaga 361

Qy     370 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 429
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     362 caagaacctggtgactggtgatcacatccccaccccacaggatctg-cccagagaaagtc 420

Qy     430 CTCGCTCGTCACCAGCAAGCTTGCGGG 456
           ||||||||||||||||||||||||||
Db     421 ctcgctcgtcaccagcaagcttgcggg 447
```

```
RESULT   3
US-08-739-484-2
; Sequence 2, Application US/08739484
; Patent No. 5917028
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    TITLE OF INVENTION:  A NOVEL HUMAN
;    TITLE OF INVENTION:  PHOSPHOPROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/739,484
```

```
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0140 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  454 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:
;      CLONE:  Concensus
US-08-739-484-2


  Query Match             24.6%;  Score 271.4;  DB 2;  Length 454;
  Best Local Similarity   99.0%;  Pred. No. 2.7e-76;
  Matches  294;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    2;

Qy    162 AGGACACGCAGGAG-AAAGAAGGTATTCTGCCTGAG-AGAGCTGAAGAGGCAAAGCTAAA 219
          ||||||||||||||| |||||||||||||||||||| |||||||||||||||||||||||
Db     61 AGGACACGCAGGAGCAAAGAAGGTATTCTGCCTGAGAAGAGCTGAAGAGGCAAAGCTAAA 120

Qy    220 GGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACT 279
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    121 GGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTTCTCATGAAGAGACT 180

Qy    280 CCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAA 339
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAA 240

Qy    340 GAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCC 399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCC 300

Qy    400 CACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 456
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGG 357


RESULT   4
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-Fls
US-08-232-463-14


  Query Match              5.8%;  Score 64.4;  DB 1;  Length 7218;
  Best Local Similarity    5.5%;  Pred. No. 7.7e-10;
  Matches   17;  Conservative  186;  Mismatches  107;  Indels    0;  Gaps    0;

Qy    114 AGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGG 173
          ::::: :::::::::::::    : ::::::::  :: :::::::::: :::: : : :::
Db   1347 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1288

Qy    174 AGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAA 233
          :::::::::: :    :   :::::::  ::::::: ::::  :::::  ::: :   ::
Db   1287 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1228

Qy    234 GCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAA 293
          :   :::: :::::   :::::    ::       : ::::::::    :::::::: :::
Db   1227 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1168

Qy    294 AGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGC 353
          :: :    ::   :::::::  : :: : ::  ::::  :::: :::::: ::: ::  :
Db   1167 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1108

Qy    354 CAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATC 413
           ::: : ::::  ::: :::::   :: ::  :: ::  : ||| | | | | |  || |
Db   1107 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCTCCCTCGACC 1048

Qy    414 TGCCCCAGAG 423
          |||  |  ||
Db   1047 TGCAGCCAAG 1038


RESULT   5
US-09-385-982-542/c
; Sequence 542, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
;  APPLICANT: ENDEGE, WILSON O., ET AL.
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS:  II
;  FILE REFERENCE: CCDNA-260XX
;  CURRENT APPLICATION NUMBER: US/09/385,982
;  CURRENT FILING DATE: 1999-08-30
;  EARLIER APPLICATION NUMBER: 09/328,111
;  EARLIER FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: 60/117,393
;  EARLIER FILING DATE: 1999-01-27
;  EARLIER APPLICATION NUMBER: 60/098,639
;  EARLIER FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 544
;  SOFTWARE: FastSEQ for Windows Version 3.0
```

```
; SEQ ID NO 542
;   LENGTH: 770
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(770)
;   OTHER INFORMATION: n = A,T,C or G
US-09-385-982-542


  Query Match             3.4%;  Score 38;  DB 4;  Length 770;
  Best Local Similarity   62.8%;  Pred. No. 0.04;
  Matches   59;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy   1008 CATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAAC 1067
           ||| ||||| ||  | ||  | |||        |||| ||||| ||  ||| |  | ||
Db     94 CATAACTTAATAAATATTGTATTTTAGATATATTCTCTAATAAAATTCAGAATTCTAAAA 35

Qy   1068 ATTATTCACAAAAAAAAAAAAAAAAAAGGGCGGCG 1101
           |  |   | ||||||||||||||||||  |  ||
Db     34 AAAAAAAAAAAAAAAAAAAAAAAAAGTACCTCG 1


RESULT   6
US-08-123-934A-5
; Sequence 5, Application US/08123934A
; Patent No. 6291206
;  GENERAL INFORMATION:
;    APPLICANT:  WOZNEY, John
;    APPLICANT:  CELESTE, Anthony J.
;    APPLICANT:  THIES, R. Scott
;    APPLICANT:  YAMAJI, No. 6291206oru
;    TITLE OF INVENTION:  RECEPTOR PROTEINS
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute Inc.- Legal Affairs
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/123,934A
;      FILING DATE:  17-SEP-1993
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAZAR, Steven R
;      REGISTRATION NUMBER:  32,618
;      REFERENCE/DOCKET NUMBER:  5203
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617 876 1170
;      TELEFAX:  617 876 5851
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3238 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  CFK1-10a
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  474..2000
US-08-123-934A-5


  Query Match             3.4%;  Score 37.4;  DB 4;  Length 3238;
  Best Local Similarity   52.2%;  Pred. No. 0.16;
  Matches   83;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy    935 ATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTG 994
          ||  || |  |  | | ||    | || ||||| | |||   |||  | || ||||  |
Db   3053 ATTACATGCATTTAAACTCTGCCAGAAAAAAAAATAACTATTTTGTTTTAATCTACTTTT 3112

Qy    995 CCCCCTTAGCCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATA 1054
              ||||     |||  |    ||    || | | |  ||  |     ||| || | |
Db   3113 TGTATTTAGTAGTTATTTGTATAAATTAAATAAACTGTTTTCAAGTCAAAAAAAAAAAAA 3172

Qy   1055 AAGATTGATGAACATTATTCACAAAAAAAAAAAAAAAAA 1093
          || |   |  || |  |   | ||||||||||||||| |
Db   3173 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATA 3211


RESULT   7
PCT-US94-10080-5
; Sequence 5, Application PC/TUS9410080
;  GENERAL INFORMATION:
;    APPLICANT:  GENETICS INSTITUTE, INC.
;    TITLE OF INVENTION:  RECEPTOR PROTEINS
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute Inc.- Legal Affairs
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/10080
;      FILING DATE:  HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/123,934
;      FILING DATE:  17-SEP-1993
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAZAR, Steven R
;      REGISTRATION NUMBER:  32,618
;      REFERENCE/DOCKET NUMBER:  5203-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8260
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3238 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  CFK1-10a
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  474..2000
PCT-US94-10080-5


  Query Match             3.4%;  Score 37.4;  DB 5;  Length 3238;
  Best Local Similarity   52.2%;  Pred. No. 0.16;
  Matches   83;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy    935 ATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTG 994
          ||  || |  |  | | ||    | || ||||| | |||   |||  | || ||||  |
Db   3053 ATTACATGCATTTAAACTCTGCCAGAAAAAAAAATAACTATTTTGTTTTAATCTACTTTT 3112
```

```
Qy    995 CCCCCTTAGCCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATA 1054
              ||||     |||  |    ||    || | | |  ||  |     ||| || | |
Db   3113 TGTATTTAGTAGTTATTTGTATAAATTAAATAAACTGTTTTCAAGTCAAAAAAAAAAAAA 3172

Qy   1055 AAGATTGATGAACATTATTCACAAAAAAAAAAAAAAAAA 1093
          || |   |  || |  |   | ||||||||||||||| |
Db   3173 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATA 3211


RESULT   8
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17


  Query Match             3.4%;  Score 37;  DB 4;  Length 289;
  Best Local Similarity   7.9%;  Pred. No. 0.043;
  Matches   18;  Conservative   94;  Mismatches  117;  Indels    0;  Gaps    0;

Qy    117 AACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGA 176
          :|  : :|: :|: :  :  : ::: : :   :  :: : : :   ::: : ::: : :
Db     45 rarcrararurgrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrs 104

Qy    177 AAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCC 236
          : : ::: :  :     :: : ::  : ::: : : :   : : :   :    :  : :
Db    105 rnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrs 164

Qy    237 TAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGT 296
            : ::: : ::  : :::    ::    :  : : : : ::    : : : ::: :|:
Db    165 rnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrcrarg 224

Qy    297 ACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAA 345
          :| : |:|  : :|:|  : :     :  : :|  | |  || || ||
Db    225 rcrurgrcrgrurararcrurcrururgrcrgrcruaaaaaaaaaaaaa 273


RESULT   9
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17


  Query Match             3.4%;  Score 37;  DB 4;  Length 289;
  Best Local Similarity   7.9%;  Pred. No. 0.043;
  Matches   18;  Conservative   94;  Mismatches  117;  Indels    0;  Gaps    0;

Qy    117 AACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGA 176
          :|  : :|: :|: :  :  : ::: : :   :  :: : : :   ::: : ::: : :
Db     45 rarcrararurgrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrs 104

Qy    177 AAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCC 236
          : : ::: :  :     :: : ::  : ::: : : :   : : :   :    :  : :
Db    105 rnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrs 164

Qy    237 TAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGT 296
            : ::: : ::  : :::    ::    :  : : : : ::    : : : ::: :|:
Db    165 rnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrcrarg 224

Qy    297 ACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAA 345
          :| : |:|  : :|:|  : :     :  : :|  | |  || || ||
Db    225 rcrurgrcrgrurararcrurcrururgrcrgrcruaaaaaaaaaaaaa 273


RESULT  10
US-08-781-891-208/c
; Sequence 208, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,891
;      FILING DATE:  27-DEC-1996
```

```
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090620tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  240052.419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  208:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16442 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-781-891-208


  Query Match             3.4%;  Score 37;  DB 3;  Length 16442;
  Best Local Similarity   47.3%;  Pred. No. 0.65;
  Matches  112;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy     113 CAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAG 172
           ||| |  | || || ||| |      |||| ||   |  ||| || |||||||  ||||
Db   16393 CAGGAGGAGGAGGAGGAGCAGGAGCAGGAGCAGGAGGAGGAGCAGGAGCAGGAGGAGCAG 16334

Qy     173 GAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCA 232
           ||| | || |     | |   |||   |  || ||||   |||   |||   |  |
Db   16333 GAGGAGGAGGAGGAGCAGGAGGAGCAGGAGGAGGAGGAGGAGCAGGAGGAGGAGCAGGAG 16274

Qy     233 AGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAA 292
             |   |  |||  |  |     ||          | ||   |   || | | | |
Db   16273 CAGGAGGAGGAGGAGCAGGAGGAGCAGGAGGAGGAGGAGCAGGAGCAGGAGGAGGAGGAG 16214

Qy     293 AAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAG 349
           || |    || |     ||  | || | |   ||    |||| |||||| ||| ||
Db   16213 GAGGAGGAGGAGTAGAAGGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 16157


RESULT  11
US-08-132-649-9
; Sequence 9, Application US/08132649
; Patent No. 5585462
;  GENERAL INFORMATION:
;    APPLICANT:  Londos, Constantine
;    APPLICANT:  Greenberg, Andrew S.
;    APPLICANT:  Kimmel, Alan R.
;    APPLICANT:  Egan, John J.
;    TITLE OF INVENTION:  CLONING OF PERILIPIN PROTEINS
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  Steuart Street Tower, One Market Plaza
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  94105-1493
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/132,649
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garrett-Wackowski, Eugenia
;      REGISTRATION NUMBER:  37,330
;      REFERENCE/DOCKET NUMBER:  15280-145-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 543-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  407 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..407
;      OTHER INFORMATION:  /note= "3' human perilipin
;      OTHER INFORMATION:  nucleotide sequences."
US-08-132-649-9


  Query Match             3.2%;  Score 35.6;  DB 1;  Length 407;
  Best Local Similarity   51.2%;  Pred. No. 0.15;
  Matches   83;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy     123 AAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAG 182
           | || ||| | | || |||  |||| | |  ||||    |||||||| ||||| | ||||
Db      31 AGGAGGAGGATCATGAGGACCAGACAGACACGGAGGGAGAGGACACGGAGGAGGAGGAAG 90

Qy     183 GTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGAC 242
             |   | |||||       |         || ||   |||             | ||| |
Db      91 AATTGGAGACTGAGGAGAACAAGTTCAGTGAGGTAGCAGCCCTGCCAGGCCCTCGAGGCC 150

Qy     243 AAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGA 284
           |  ||| |   |  |   ||||  |||||  |||||||
Db     151 TCCTGGGTGGTGTGGCACATACCCTGCAGAAGACCCTCCAGA 192


RESULT  12
US-08-767-579-9
; Sequence 9, Application US/08767579
; Patent No. 6074842
;  GENERAL INFORMATION:
;    APPLICANT:  Londos, Constantine
;    APPLICANT:  Greenberg, Andrew S.
;    APPLICANT:  Kimmel, Alan R.
;    APPLICANT:  Egan, John J.
;    TITLE OF INVENTION:  CLONING OF PERILIPIN PROTEINS
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  Steuart Street Tower, One Market Plaza
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  94105-1493
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/767,579
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garrett-Wackowski, Eugenia
;      REGISTRATION NUMBER:  37,330
;      REFERENCE/DOCKET NUMBER:  15280-145-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 543-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  407 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..407
;      OTHER INFORMATION:  /note= "3' human perilipin
;      OTHER INFORMATION:  nucleotide sequences."
US-08-767-579-9


  Query Match             3.2%;  Score 35.6;  DB 3;  Length 407;
  Best Local Similarity   51.2%;  Pred. No. 0.15;
  Matches   83;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy    123 AAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAG 182
          | || ||| | | || |||  |||| | |  ||||    |||||||| ||||| | ||||
Db     31 AGGAGGAGGATCATGAGGACCAGACAGACACGGAGGGAGAGGACACGGAGGAGGAGGAAG 90

Qy    183 GTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGAC 242
              |   | |||||       |         || ||   |||             | ||| |
Db     91 AATTGGAGACTGAGGAGAACAAGTTCAGTGAGGTAGCAGCCCTGCCAGGCCCTCGAGGCC 150

Qy    243 AAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGA 284
            |  ||| |   |  |   |||   |||||  |||||||
Db    151 TCCTGGGTGGTGTGGCACATACCCTGCAGAAGACCCTCCAGA 192


RESULT  13
US-08-132-649-5
; Sequence 5, Application US/08132649
; Patent No. 5585462
;  GENERAL INFORMATION:
;    APPLICANT:  Londos, Constantine
;    APPLICANT:  Greenberg, Andrew S.
;    APPLICANT:  Kimmel, Alan R.
;    APPLICANT:  Egan, John J.
;    TITLE OF INVENTION:  CLONING OF PERILIPIN PROTEINS
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  Steuart Street Tower, One Market Plaza
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  94105-1493
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/132,649
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garrett-Wackowski, Eugenia
;      REGISTRATION NUMBER:  37,330
;      REFERENCE/DOCKET NUMBER:  15280-145-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 543-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1265 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..1265
;      OTHER INFORMATION:  /standard_name= "HUPERI"
;      OTHER INFORMATION:  /note= "Nucleotide sequence of human Perilipin
;      OTHER INFORMATION:  cDNA."
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..1265
US-08-132-649-5


  Query Match             3.2%;  Score 35.6;  DB 1;  Length 1265;
  Best Local Similarity   51.2%;  Pred. No. 0.32;
  Matches   83;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy    123 AAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAG 182
          | || ||| | | || |||  |||| | |  ||||    |||||||| ||||| | ||||
Db    889 AGGAGGAGGATCATGAGGACCAGACAGACACGGAGGGAGAGGACACGGAGGAGGAGGAAG 948

Qy    183 GTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGAC 242
              |   | |||||       |         || ||   |||             | ||| |
Db    949 AATTGGAGACTGAGGAGAACAAGTTCAGTGAGGTAGCAGCCCTGCCAGGCCCTCGAGGCC 1008

Qy    243 AAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGA 284
            |  ||| |   |  |   |||   |||||  |||||||
Db   1009 TCCTGGGTGGTGTGGCACATACCCTGCAGAAGACCCTCCAGA 1050


RESULT  14
US-08-767-579-5
; Sequence 5, Application US/08767579
; Patent No. 6074842
;  GENERAL INFORMATION:
;    APPLICANT:  Londos, Constantine
;    APPLICANT:  Greenberg, Andrew S.
;    APPLICANT:  Kimmel, Alan R.
;    APPLICANT:  Egan, John J.
;    TITLE OF INVENTION:  CLONING OF PERILIPIN PROTEINS
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  Steuart Street Tower, One Market Plaza
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  94105-1493
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/767,579
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garrett-Wackowski, Eugenia
;      REGISTRATION NUMBER:  37,330
;      REFERENCE/DOCKET NUMBER:  15280-145-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 543-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1265 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
```

```
;       ORGANISM:  Homo sapiens
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1..1265
;       OTHER INFORMATION:  /standard_name= "HUPERI"
;       OTHER INFORMATION:  /note= "Nucleotide sequence of human Perilipin
;       OTHER INFORMATION:  cDNA."
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  3..1265
US-08-767-579-5
```

```
  Query Match             3.2%;  Score 35.6;  DB 3;  Length 1265;
  Best Local Similarity   51.2%;  Pred. No. 0.32;
  Matches   83;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy     123 AAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAG 182
           | || ||| | | || ||| |||| | |  ||||    |||||||| ||||| | ||||
Db     889 AGGAGGAGGATCATGAGGACCAGACAGACACGGAGGGAGAGGACACGGAGGAGGAGGAAG 948

Qy     183 GTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGAC 242
             |   | |||||       |        || ||   |||          | ||| |
Db     949 AATTGGAGACTGAGGAGAACAAGTTCAGTGAGGTAGCAGCCCTGCCAGGCCCTCGAGGCC 1008

Qy     243 AAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGA 284
             | ||| |   |  |   |||   |||||  |||||||
Db    1009 TCCTGGGTGGTGTGGCACATACCCTGCAGAAGACCCTCCAGA 1050
```

```
RESULT  15
US-08-459-568-3/c
; Sequence 3, Application US/08459568
; Patent No. 5811304
;  GENERAL INFORMATION:
;    APPLICANT:  Huang, Shi
;    TITLE OF INVENTION:  Retinoblastoma Protein - Interacting
;    TITLE OF INVENTION:  Zinc Finger Proteins
;    NUMBER OF SEQUENCES:  93
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/459,568
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/399,411
;      FILING DATE:  06-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 1264
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5183 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..5158
US-08-459-568-3
```

```
  Query Match             3.2%;  Score 35;  DB 1;  Length 5183;
  Best Local Similarity   57.9%;  Pred. No. 1.3;
  Matches   62;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     117 AACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGA 176
           ||  |||||| ||| |   ||  || || | ||  |||||| || | ||     |||||
Db    3223 AAGGAGAAGAGGAGGAAGATGAAGAAGAAAACGAAGAGGAGGAGGAAGATGAAGAGGAGG 3164

Qy     177 AAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCC 223
           ||||||  | | |    |     | ||| | |||| |  | | || |
Db    3163 AAGAAGAAAGTGTTGGAGGCCCGGGTGAGGCGGCAGACATCAGGGGC 3117
```

```
RESULT  16
US-08-399-411-3/c
; Sequence 3, Application US/08399411
; Patent No. 5831008
;  GENERAL INFORMATION:
;    APPLICANT:  Huang, Shi
;    TITLE OF INVENTION:  Retinoblastoma Protein - Interacting
;    TITLE OF INVENTION:  Zinc Finger Proteins
;    NUMBER OF SEQUENCES:  93
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/399,411
;      FILING DATE:  06-MAR-1995
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 1264
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5183 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..5158
US-08-399-411-3
```

```
  Query Match             3.2%;  Score 35;  DB 2;  Length 5183;
  Best Local Similarity   57.9%;  Pred. No. 1.3;
  Matches   62;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     117 AACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGA 176
           ||  |||||| ||| |   ||  || || | ||  |||||| || | ||     |||||
Db    3223 AAGGAGAAGAGGAGGAAGATGAAGAAGAAAACGAAGAGGAGGAGGAAGATGAAGAGGAGG 3164

Qy     177 AAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCC 223
           ||||||  | | |    |     | ||| | |||| |  | | || |
Db    3163 AAGAAGAAAGTGTTGGAGGCCCGGGTGAGGCGGCAGACATCAGGGGC 3117
```

```
RESULT  17
US-08-516-859A-3/c
; Sequence 3, Application US/08516859A
; Patent No. 6069231
;  GENERAL INFORMATION:
;    APPLICANT:  Huang, Shi
;    TITLE OF INVENTION:  Retinoblastoma Protein - Interacting
;    TITLE OF INVENTION:  Zinc Finger Proteins
;    NUMBER OF SEQUENCES:  106
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/516,859A
;      FILING DATE:  18-AUG-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/399,411
;      FILING DATE:  06-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/292,683
;      FILING DATE:  18-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 1776
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5868 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  121..5278
US-08-516-859A-3


  Query Match             3.2%;  Score 35;  DB 3;  Length 5868;
  Best Local Similarity   57.9%;  Pred. No. 1.4;
  Matches   62;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     117 AACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGA 176
           ||  |||||| ||| |   ||  || || | ||  |||||| || | ||     |||||
Db    3343 AAGGAGAAGAGGAGGAAGATGAAGAAGAAAACGAAGAGGAGGAGGAAGATGAAGAGGAGG 3284

Qy     177 AAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCC 223
           ||||||  | | |    |     | ||| | |||| |  | | || |
Db    3283 AAGAAGAAAGTGTTGGAGGCCCGGGTGAGGCGGCAGACATCAGGGGC 3237


RESULT  18
US-08-557-892-5/c
; Sequence 5, Application US/08557892
; Patent No. 5968818
;  GENERAL INFORMATION:
;    APPLICANT:  MacPhee, Colin Houston
;    APPLICANT:  Tew, David Graham
;    APPLICANT:  Southan, Christopher Donald
;    APPLICANT:  Hickey, Dierdre Mary Bernadette
;    APPLICANT:  Gloger, Israel Simon
;    APPLICANT:  Lawrence, Geoffrey Mark Prouse
;    APPLICANT:  Rice, Simon Quentyn John
;    TITLE OF INVENTION:  Compounds
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/557,892
;      FILING DATE:  14 No. 5968818ember 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/387,858
;      FILING DATE:  14 No. 5968818ember 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dustman, Wayne J.
;      REGISTRATION NUMBER:  33,870
;      REFERENCE/DOCKET NUMBER:  P30693C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5023
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  420 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-557-892-5


  Query Match             3.1%;  Score 34.4;  DB 2;  Length 420;
  Best Local Similarity   58.3%;  Pred. No. 0.37;
  Matches   56;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy     997 CCCTTAGCCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAA 1056
           |||| |||| ||  |  |||  | | || |   ||| |  |  |   | ||||  |||||
Db     415 CCCTCAGCCNCCNCTTTTTANCAACNCTGAATATTTCCAATATCCTGCTAATATCATAAA 356

Qy    1057 GATTGATGAACATTATTCACAAAAAAAAAAAAAAAA 1092
             ||  |  ||   || ||||   | ||| ||| |||
Db     355 AATGAAAAAATGCTACTCACCTGATAAAGAAAGAAA 320


RESULT  19
US-08-387-858A-5/c
; Sequence 5, Application US/08387858A
; Patent No. 5981252
;  GENERAL INFORMATION:
;    APPLICANT:  MacPhee, Colin Houston
;    APPLICANT:  Tew, David Graham
;    APPLICANT:  Southan, Christopher Donald
;    APPLICANT:  Hickey, Dierdre Mary Bernadette
;    APPLICANT:  Gloger, Israel Simon
;    APPLICANT:  Lawrence, Geoffrey Mark Prouse
;    APPLICANT:  Rice, Simon Quentyn John
;    TITLE OF INVENTION:  Compounds
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
```

```
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/387,858A
;      FILING DATE:  24 February 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB94/01374
;      FILING DATE:  24 June 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dustman, Wayne J.
;      REGISTRATION NUMBER:  33,870
;      REFERENCE/DOCKET NUMBER:  P30693
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5023
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  420 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-387-858A-5


  Query Match             3.1%;  Score 34.4;  DB 2;  Length 420;
  Best Local Similarity   58.3%;  Pred. No. 0.37;
  Matches   56;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy     997 CCCTTAGCCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAA 1056
           |||| |||| ||  |  |||  | | || |   ||| |  |  |   | ||||  |||||
Db     415 CCCTCAGCCNCCNCTTTTTANCAACNCTGAATATTTCCAATATCCTGCTAATATCATAAA 356

Qy    1057 GATTGATGAACATTATTCACAAAAAAAAAAAAAAAA 1092
             ||  |  ||   || ||||   | ||| ||| |||
Db     355 AATGAAAAAATGCTACTCACCTGATAAAGAAAGAAA 320


RESULT  20
US-09-294-384B-5/c
; Sequence 5, Application US/09294384B
; Patent No. 6177257
;  GENERAL INFORMATION:
;    APPLICANT:  MacPhee, Colin Houston
;    APPLICANT:  Tew, David Graham
;    APPLICANT:  Southan, Christopher Donald
;    APPLICANT:  Hickey, Dierdre Mary Bernadette
;    APPLICANT:  Gloger, Israel Simon
;    APPLICANT:  Lawrence, Geoffrey Mark Prouse
;    APPLICANT:  Rice, Simon Quentyn John
;    TITLE OF INVENTION:  Compounds
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/294,384B
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/387,858
;      FILING DATE:  24 February 1995
;      APPLICATION NUMBER:  PCT/GB94/01374
;      FILING DATE:  24 June 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dustman, Wayne J.
;      REGISTRATION NUMBER:  33,870
;      REFERENCE/DOCKET NUMBER:  P30693
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5023
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  420 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-09-294-384B-5


  Query Match             3.1%;  Score 34.4;  DB 4;  Length 420;
  Best Local Similarity   58.3%;  Pred. No. 0.37;
  Matches   56;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy     997 CCCTTAGCCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAA 1056
           |||| |||| ||  |  |||  | | || |   ||| |  |  |   | ||||  |||||
Db     415 CCCTCAGCCNCCNCTTTTTANCAACNCTGAATATTTCCAATATCCTGCTAATATCATAAA 356

Qy    1057 GATTGATGAACATTATTCACAAAAAAAAAAAAAAAA 1092
             ||  |  ||   || ||||   | ||| ||| |||
Db     355 AATGAAAAAATGCTACTCACCTGATAAAGAAAGAAA 320


RESULT  21
US-08-145-705A-1/c
; Sequence 1, Application US/08145705A
; Patent No. 5489513
;  GENERAL INFORMATION:
;    APPLICANT:  Springer, Wolfgang; Plempel,Manfred;
;    APPLICANT:  L bberding, Antonius
;    TITLE OF INVENTION:  SPECIFIC GENE PROBES AND
;    TITLE OF INVENTION:  PROCESSES FOR THE DIAGNOSTIC
;    TITLE OF INVENTION:  INVESTIGATION OF CANDIDA
;    TITLE OF INVENTION:  ALBICANS
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SPRUNG HORN KRAMER & WOODS
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.4 MB storage
;      COMPUTER:  NEC PowerMate 1 Plus
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/145,705A
;      FILING DATE:  October 28, 1993
;      CLASSIFICATION:  536
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  German P 42 36 708.5
;      FILING DATE:  October 30, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  Bayer 8885-KGB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332-1700
;      TELEFAX:  (914) 332-1844
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2233 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Candida albicans
US-08-145-705A-1


  Query Match             3.1%;  Score 34.2;  DB 1;  Length 2233;
  Best Local Similarity   55.5%;  Pred. No. 1.3;
  Matches   66;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy    104 GCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAG 163
          |||| |    || || |||| |   ||||    || ||||||  ||    || ||    ||
Db    258 GCCAAGAAGAAGGAAGAAGAGGCCAAGAAGAAGGCAGAGGAGGCCAAGAAGAAGGCAGAG 199

Qy    164 GACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGC 222
          ||  |  || |||| | ||            |||  || || |||||  |||   |||||
Db    198 GAGGCCAAGAAGAAGGCAGAGGAGGCCAAGAAGAAGGCAGAGGAGGCCAAGAAGAAGGC 140


RESULT  22
US-09-156-809-3
; Sequence 3, Application US/09156809
; Patent No. 6316262
; GENERAL INFORMATION:
;  APPLICANT: Huisman, Gjalt W.
;  APPLICANT: Skraly, Frank A.
;  APPLICANT: Martin, David P.
;  APPLICANT: Peoples, Oliver P.
;  TITLE OF INVENTION: Biological Systems for the Manufacture of
;  TITLE OF INVENTION: Polyhydroxyalkanoate Polymers Containing 4-Hydroxyacids
;  FILE REFERENCE: MBX 017 PTO
;  CURRENT APPLICATION NUMBER: US/09/156,809
;  CURRENT FILING DATE: 1998-09-18
;  EARLIER APPLICATION NUMBER: 60/059,373
;  EARLIER FILING DATE: 1997-09-18
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 1289
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: orzf gene from
;   OTHER INFORMATION: C. kluyveri
US-09-156-809-3


  Query Match             3.1%;  Score 34;  DB 4;  Length 1289;
  Best Local Similarity   49.4%;  Pred. No. 1.1;
  Matches   88;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy    912 TGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGA 971
          || ||    | ||   | | |   ||| || | |||| ||   |   |  ||||||
Db    548 tgttagaattgcagcctcctaaattgggagatgtagaaaaagccataggagaaaactgtg 607

Qy    972 CTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTATCTCTTACAGTT 1031
          | |  ||   | ||  |  ||   | | | ||| |  | ||  |  |||  |  |
Db    608 catctttaattgaagatggagctactcttcagcttggaataggtgctataccagatgcgg 667

Qy   1032 TGCTCTGGCTCTCAAATAATATAAAGATTGATGAACATTATTCACAAAAAAAAAAAAA 1089
          | || |   ||| ||| || | ||||| |   | |    ||||  | |  | |  | |
Db    668 tacttttattcttaaagaacaaaaagaatttaggaatacattctgagatgatatcaga 725


RESULT  23
US-08-931-999-4
; Sequence 4, Application US/08931999
; Patent No. 6043219
;  GENERAL INFORMATION:
;    APPLICANT:  Iandolo, John J.
;    APPLICANT:  Crupper, Scott S.
;    TITLE OF INVENTION:  Broad Spectrum Chemotherapeutic Peptide
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hovey, Williams, Timmons & Collins
;      STREET:  2405 Grand Boulevard, Suite 400
;      CITY:  Kansas City
;      STATE:  Missouri
;      COUNTRY:  U.S.A.
;      ZIP:  64108
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/931,999
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/710,561
;      FILING DATE:  19-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Collins, John M.
;      REGISTRATION NUMBER:  26,262
;      REFERENCE/DOCKET NUMBER:  25043-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  816/474-9050
;      TELEFAX:  816/474-9057
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6755 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Staphylococcus aureus
;      STRAIN:  UT0007
US-08-931-999-4


  Query Match             3.1%;  Score 34;  DB 3;  Length 6755;
  Best Local Similarity   44.8%;  Pred. No. 3.2;
  Matches  130;  Conservative    0;  Mismatches  160;  Indels    0;  Gaps    0;

Qy    120 AAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAG 179
          || |  |  ||||    |  ||  | |  ||  |  |||   |  | | ||   | |||
Db   6378 AAAAGAAGAAGAAAAGGGAAGAAAAAAGAGGAAAACAGAGAGAACAAAAGCCAAAAAAAC 6437

Qy    180 AAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAG 239
          || | |        || |||||  ||      | || |||||  |||| |  || | | |
Db   6438 AACGAAGCGGAGAAGAAAGAGCAAAAACAAACACGCAAAAGGGCAAAAAAGCACGACCAA 6497

Qy    240 GACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACT 299
```

```
            |||  |    | || |       |  |  |    || | | |||  |  ||   |
Db   6498 ACAAAAGACAAACAAGCACAACAAAGCGAACAACACAACCCAAAAAAAAGAGAACAGAAA 6557

Qy    300 TTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTG 359
            |  || ||||  |  | |       || ||| |  |  ||  |  |    | |||
Db   6558 AACAAACACGAGAACAACAAAGAAAGGGAGGCAAAAGAACAAAACAAAAAACACGAAGAA 6617

Qy    360 CAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAG 409
          || |   | | ||  |       |  |   |  ||| |||| ||||  ||
Db   6618 CACGGGGAAAAAAAGAGGACACAAAGGAAAACAACACCCCCCCCCCCAAG 6667


RESULT  24
US-09-040-984-16/c
; Sequence 16, Application US/09040984
; Patent No. 6210883
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Wang, TongTong
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR DIAGNOSIS
;    TITLE OF INVENTION:  OF LUNG CANCER
;    NUMBER OF SEQUENCES:  86
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/040,984
;      FILING DATE:  18-MAR-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.456
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-622-4900
;      TELEFAX:  206-282-6031
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  669 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-040-984-16


  Query Match             3.1%;  Score 33.8;  DB 4;  Length 669;
  Best Local Similarity   49.7%;  Pred. No. 0.79;
  Matches   86;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy      19 GCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCC 78
           ||||   |  || |  |   | |   |||  |||| | || |||||  ||| | | ||
Db     183 GCAGAATACCTTCTTTCTCCTGCGTGTCCTGCTTCTCCTCGCCGGTCTCCTCCGCAGGGT 124

Qy      79 CCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGC 138
             ||     ||     ||  || |  | |  | ||  ||     ||     ||   || |
Db     123 TCTCTTCTTCTTGTTTCTGGGACATGGCGGGACCGGGACTGTGGAGTGTAAGGGGCCCGG 64

Qy     139 GGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGC 191
           | |||  ||||| || | ||||  |||   | |   |  | ||  |   ||||
Db      63 GAAGGCAACCGGAGAAGGGAAGGGGGAGGGGAAACGGGGACAACCTGCGCTGC 11


RESULT  25
```

```
US-09-123-912-16/c
; Sequence 16, Application US/09123912A
; Patent No. 6312695
; GENERAL INFORMATION:
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Wang, Tongtong
;  TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
;  FILE REFERENCE:  210121.455C1
;  CURRENT APPLICATION NUMBER: US/09/123,912A
;  CURRENT FILING DATE:  1998-07-27
;  PRIOR APPLICATION NUMBER:  09/040,802
;  PRIOR FILING DATE:  1998-03-18
;  NUMBER OF SEQ ID NOS:  114
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 16
;   LENGTH: 669
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (299)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (354)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (483)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (555)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (571)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (573)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (577)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (642)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (651)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (662)
;   OTHER INFORMATION: Where n is a, c, g or t
;   NAME/KEY: modified_base
;   LOCATION: (667)
;   OTHER INFORMATION: Where n is a, c, g or t
US-09-123-912-16


  Query Match             3.1%;  Score 33.8;  DB 4;  Length 669;
  Best Local Similarity   49.7%;  Pred. No. 0.79;
  Matches   86;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy      19 GCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCC 78
           ||||   |  || |  |   | |   |||  |||| | || |||||  ||| | | ||
Db     183 GCAGAATACCTTCTTTCTCCTGCGTGTCCTGCTTCTCCTCGCCGGTCTCCTCCGCAGGGT 124

Qy      79 CCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGC 138
             ||     ||     ||  || |  | |  | ||  ||     ||     ||   || |
Db     123 TCTCTTCTTCTTGTTTCTGGGACATGGCGGGACCGGGACTGTGGAGTGTAAGGGGCCCGG 64

Qy     139 GGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGC 191
           | |||  ||||| || | ||||  |||   | |   |  | ||  |   ||||
Db      63 GAAGGCAACCGGAGAAGGGAAGGGGGAGGGGAAACGGGGACAACCTGCGCTGC 11


RESULT  26
```

```
US-08-937-540-3
; Sequence 3, Application US/08937540
; Patent No. 5891697
;  GENERAL INFORMATION:
;    APPLICANT:  Croteau, Rodney B
;    APPLICANT:  Wise, Mitchell L
;    APPLICANT:  Savage, Thomas J
;    APPLICANT:  Katahira, Eva J
;    TITLE OF INVENTION:  Monoterpene Synthases from Common Sage
;    TITLE OF INVENTION:  (Salvia officinalis)
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHRISTENSEN, O'CONNOR, JOHNSON & KINDNESS
;      STREET:  1420 FIFTH AVENUE
;      CITY:  SEATTLE
;      STATE:  WASHINGTON
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/937,540
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shelton, Dennis K
;      REGISTRATION NUMBER:  26,997
;      REFERENCE/DOCKET NUMBER:  WSUR111254
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206 695 1718
;      TELEFAX:  206 224 0779
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1968 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Salvia officinalis
;    IMMEDIATE SOURCE:
;      CLONE:  1,8-cineole synthase
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14..1788
US-08-937-540-3


  Query Match             3.0%;  Score 33.4;  DB 2;  Length 1968;
  Best Local Similarity   68.7%;  Pred. No. 2.2;
  Matches   46;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy    1027 CAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACATTATTCACAAAAAAAAAA 1086
           || || ||||  | ||  ||  |||| || | || | | || ||   | ||||||||||
Db    1899 CATTTCGCTCATGTTCATAATAAATAATAAAACTGTTAATCAATAACAAAAAAAAAAAA 1958

Qy    1087 AAAAAAA 1093
           |||||||
Db    1959 AAAAAAA 1965


RESULT  27
US-08-781-891-209/c
; Sequence 209, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,891
;      FILING DATE:  27-DEC-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090620tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  240052.419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  209:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  51259 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-781-891-209


  Query Match             3.0%;  Score 33.4;  DB 3;  Length 51259;
  Best Local Similarity   55.7%;  Pred. No. 19;
  Matches   64;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy   114 AGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGG 173
         || | || || || ||| |      |||||||   |   ||||| || |||| | | |||
Db   174 AGGAGCAGGAGGAGGAGGAGGAGCAGGAGGAGGAGGAGAAGGAGGAGAAGGAGAAGGAGG 115

Qy   174 AGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATA 228
         |||| || ||      |   |||   |  || ||||   |||   ||    || |
Db   114 AGAAGGAGGGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAGGAGCAGGAACA 60


RESULT  28
US-08-770-379-20/c
; Sequence 20, Application US/08770379
; Patent No. 5849564
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
;    TITLE OF INVENTION:  HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,379
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  52342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32207 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-770-379-20


  Query Match             3.0%;  Score 33.2;  DB 2;  Length 32207;
  Best Local Similarity   59.6%;  Pred. No. 16;
  Matches   56;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy      20 CAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCC 79
           || | |   ||  ||||   ||||| |   |  ||| || || ||| ||  ||    |||
Db  32101 CATCCCCCCTTTCCCCCCCATCCCCCCTTTCCCCCCATCCCCCGTTTCCCCCCATCCCCC 32042

Qy      80 CTTACACTCCACAGTCCCGGTCCCGCCATGTCCC 113
           ||| | | |||    |||  |||| ||||  |||
Db  32041 CTTTCCCCCCATTCCCCCTTTCCCCCCATCCCCC 32008


RESULT  29
US-08-757-669A-20/c
; Sequence 20, Application US/08757669A
; Patent No. 6183751
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
;    TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/757,669A
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  45185-F
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32207 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-757-669A-20


  Query Match             3.0%;  Score 33.2;  DB 4;  Length 32207;
  Best Local Similarity   59.6%;  Pred. No. 16;
  Matches   56;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy      20 CAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCC 79
           || | |   ||  ||||   ||||| |   |  ||| || || ||| ||  ||    |||
Db  32101 CATCCCCCCTTTCCCCCCCATCCCCCCTTTCCCCCCATCCCCCGTTTCCCCCCATCCCCC 32042

Qy      80 CTTACACTCCACAGTCCCGGTCCCGCCATGTCCC 113
           ||| | | |||    |||  |||| ||||  |||
Db  32041 CTTTCCCCCCATTCCCCCTTTCCCCCCATCCCCC 32008


RESULT  30
US-09-230-371A-20/c
; Sequence 20, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
;  APPLICANT:  Chang, Yuan
;  APPLICANT:  Bohenzky, Roy A
;  APPLICANT:  Russo, James J
;  APPLICANT:  Edelman, Isidore S
;  APPLICANT:  Moore, Patrick S
;  TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE:  45185-G-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/230,371A
;  CURRENT FILING DATE:  1999-11-17
;  PRIOR APPLICATION NUMBER:  PCT/US97/13346
;  PRIOR FILING DATE:  1997-07-22
;  NUMBER OF SEQ ID NOS:  30
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 32207
;   TYPE: DNA
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20


  Query Match             3.0%;  Score 33.2;  DB 4;  Length 32207;
  Best Local Similarity   59.6%;  Pred. No. 16;
  Matches   56;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy      20 CAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCC 79
           || | |   ||  ||||   ||||| |   |  ||| || || ||| ||  ||    |||
Db  32101 CATCCCCCCTTTCCCCCCCATCCCCCCTTTCCCCCCATCCCCCGTTTCCCCCCATCCCCC 32042

Qy      80 CTTACACTCCACAGTCCCGGTCCCGCCATGTCCC 113
           ||| | | |||    |||  |||| ||||  |||
Db  32041 CTTTCCCCCCATTCCCCCTTTCCCCCCATCCCCC 32008


RESULT  31
US-09-286-529-6
; Sequence 6, Application US/09286529
; Patent No. 6297367
; GENERAL INFORMATION:
;  APPLICANT:  Catherine Tribouley
;  TITLE OF INVENTION:  NEW MEMBERS OF TNF AND TNFR FAMILIES
;  FILE REFERENCE:  1408.003/200130.439C1
;  CURRENT APPLICATION NUMBER: US/09/286,529
;  CURRENT FILING DATE:  1999-04-05
;  NUMBER OF SEQ ID NOS:  25
```

```
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 1284
;   TYPE: DNA
;   ORGANISM: human
US-09-286-529-6


  Query Match             3.0%;  Score 33;  DB 4;  Length 1284;
  Best Local Similarity   54.5%;  Pred. No. 2.2;
  Matches   66;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy     973 TTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTATCTCTTACAGTTT 1032
           ||| |    | ||||| | ||| ||   |  |  |   |   |||    | |      ||
Db    1161 ttggtgcattgaaactgctgtgacctacttacaccatgtctgtagctattttcctccctt 1220

Qy    1033 GCTCTGGCTCTCAAATAATATAAAGATTGATGAACATTATTCACAAAAAAAAAAAAAAAA 1092
            |||||   ||| || || | | |   | |   | | ||   | ||||||||||||||||
Db    1221 tctctgtacctctaagaagaaagaatctaactgaaaatacaaaaaaaaaaaaaaaaaaaa 1280

Qy    1093 A 1093
           |
Db    1281 a 1281


RESULT  32
US-08-504-459-5
; Sequence 5, Application US/08504459
; Patent No. 5922563
;  GENERAL INFORMATION:
;    APPLICANT:  Alderete, John F.
;    TITLE OF INVENTION:  Adhesin Genes and Proteins Involved in
;    TITLE OF INVENTION:  Trichomonas Vaginalis Cytoadherence
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  US
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/504,459
;      FILING DATE:  Concurrently herewith
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wilson, Mark B.
;      REGISTRATION NUMBER:  37,259
;      REFERENCE/DOCKET NUMBER:  UTSK:273/WIM
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  512/418-3000
;      TELEFAX:  512/474-7577
;      TELEX:  N/A
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1804 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14..1714
US-08-504-459-5


  Query Match             3.0%;  Score 33;  DB 2;  Length 1804;
  Best Local Similarity   60.7%;  Pred. No. 2.7;
  Matches   54;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy    1005 CTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATG 1064
           ||| | ||   ||| | |   ||  |||    || || | ||| || | ||| |   |
Db    1709 CTCTACTAAGCAGTTTTTAAAACTCTTTCAATTGTCTTTGAAAAAAAAAAAAAAAAAAAA 1768

Qy    1065 AACATTATTCACAAAAAAAAAAAAAAAA 1093
           || |  |   | ||||||||||||||||
Db    1769 AAAAAAAAAAAAAAAAAAAAAAAAAAAA 1797


RESULT  33
US-08-676-967-2
; Sequence 2, Application US/08676967
; Patent No. 5747317
;  GENERAL INFORMATION:
;    APPLICANT:  COLLINS, KATHLEEN
;    TITLE OF INVENTION:  Human Telomerase
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Science & Technology Law Group
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/676,967
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UCB96-055
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415)343-4342
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2277 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-676-967-2


  Query Match             3.0%;  Score 33;  DB 1;  Length 2277;
  Best Local Similarity   44.2%;  Pred. No. 3.2;
  Matches   42;  Conservative   18;  Mismatches   35;  Indels    0;  Gaps    0;

Qy     120 AAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAG 179
           |:||:||:|| |     | :||:||:| :| :|| ||  |  | ||:    | ||: |:|
Db     671 ARGARGARGAYATGGARGARGARGARAAYGAYGAYGAYGAYGAYGAYGAYGARGARG 730

Qy     180 AAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAG 214
           | ||  |  |   :||   :| :||:||:| : |:
Db     731 AYGGNGTNTTYGAYGAYGARGAYGARGARGARGAR 765


RESULT  34
US-08-676-974-2
; Sequence 2, Application US/08676974
; Patent No. 5770422
;  GENERAL INFORMATION:
;    APPLICANT:  COLLINS, KATHLEEN
;    TITLE OF INVENTION:  Human Telomerase
```

```
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Science & Technology Law Group
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/676,974
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UCB96-055
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415)343-4342
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2277 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-676-974-2


  Query Match             3.0%;  Score 33;  DB 1;  Length 2277;
  Best Local Similarity   44.2%;  Pred. No. 3.2;
  Matches   42;  Conservative   18;  Mismatches   35;  Indels    0;  Gaps    0;

Qy    120 AAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAG 179
          |:||:||:|| |     | :||:||:| :| :|| ||  |  | ||:    | ||: |:|
Db    671 ARGARGARGAYATGGARGARGARGARAAYGAYGAYGAYGAYGAYGAYGAYGARGARG 730

Qy    180 AAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAG 214
          | ||  |  |   :||   :| :||:||:| : |:
Db    731 AYGGNGTNTTYGAYGAYGARGAYGARGARGARGAR 765


RESULT  35
US-09-098-487-2
; Sequence 2, Application US/09098487
; Patent No. 5917025
;  GENERAL INFORMATION:
;    APPLICANT:  COLLINS, Kathleen
;    TITLE OF INVENTION:  Human Telomerase
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Science & Technology Law Group
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/098,487
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman Ph.D., Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UCB96-055
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415)343-4342
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2277 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-098-487-2


  Query Match             3.0%;  Score 33;  DB 2;  Length 2277;
  Best Local Similarity   44.2%;  Pred. No. 3.2;
  Matches   42;  Conservative   18;  Mismatches   35;  Indels    0;  Gaps    0;

Qy    120 AAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAG 179
          |:||:||:|| |     | :||:||:| :| :|| ||  |  | ||:    | ||: |:|
Db    671 ARGARGARGAYATGGARGARGARGARAAYGAYGAYGAYGAYGAYGAYGAYGARGARG 730

Qy    180 AAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAG 214
          | ||  |  |   :||   :| :||:||:| : |:
Db    731 AYGGNGTNTTYGAYGAYGARGAYGARGARGARGAR 765


RESULT  36
US-08-131-365B-41
; Sequence 41, Application US/08131365B
; Patent No. 5527690
;  GENERAL INFORMATION:
;    APPLICANT:  Brown, Michael S.
;    APPLICANT:  Briggs, Michael R.
;    APPLICANT:  Wang, Xiaodong
;    APPLICANT:  Goldstein, Joseph L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS RELATING
;    TITLE OF INVENTION:  TO STEROL REGULATORY ELEMENT BINDING
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/131,365B
;      FILING DATE:  01-OCT-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UTSD:372/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  547 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
;    FEATURE:
```

```
;      NAME/KEY:  CDS
;      LOCATION:  1..111
US-08-131-365B-41


  Query Match             3.0%;  Score 32.8;  DB 1;  Length 547;
  Best Local Similarity   61.9%;  Pred. No. 1.4;
  Matches   52;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy   1010 TTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACAT 1069
          ||  ||  | | | || | |  |||  || || |    || | | |||    || |||||
Db    448 TTTTTTTTTTTTTTTTTCTGAATGCAATGACTGTTTTTTACTCTTAAGGAAAATAAACAT 507

Qy   1070 TATTCACAAAAAAAAAAAAAAAAA 1093
               || | |||  ||||||||||||
Db    508 CTTTTAGAAACAAAAAAAAAAAA 531


RESULT  37
US-08-668-123-41
; Sequence 41, Application US/08668123
; Patent No. 5891631
;  GENERAL INFORMATION:
;    APPLICANT:  Brown, Michael S.
;    APPLICANT:  Briggs, Michael R.
;    APPLICANT:  Wang, Xiaodong
;    APPLICANT:  Goldstein, Joseph L.
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS RELATING
;    TITLE OF INVENTION:  TO STEROL REGULATORY ELEMENT BINDING
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/668,123
;      FILING DATE:  14-JUN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/131,365
;      FILING DATE:  01-OCT-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UTSD:372/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  41:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  547 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..111
US-08-668-123-41


  Query Match             3.0%;  Score 32.8;  DB 2;  Length 547;
  Best Local Similarity   61.9%;  Pred. No. 1.4;
  Matches   52;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy   1010 TTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACAT 1069
          ||  ||  | | | || | |  |||  || || |    || | | |||    || |||||
Db    448 TTTTTTTTTTTTTTTTTCTGAATGCAATGACTGTTTTTTACTCTTAAGGAAAATAAACAT 507

Qy   1070 TATTCACAAAAAAAAAAAAAAAAA 1093
               || | |||  ||||||||||||
Db    508 CTTTTAGAAACAAAAAAAAAAAA 531


RESULT  38
US-08-630-915A-23
; Sequence 23, Application US/08630915A
; Patent No. 6309820
;  GENERAL INFORMATION:
;    APPLICANT:  SPARKS, Andrew B.
;    APPLICANT:  HOFFMAN, No. 6309820h
;    APPLICANT:  KAY, Brian K.
;    APPLICANT:  FOWLKES, Dana M.
;    APPLICANT:  McCONNELL, Stephen J.
;    TITLE OF INVENTION:  POLYPEPTIDES HAVING A FUNCTIONAL
;    TITLE OF INVENTION:  DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
;    TITLE OF INVENTION:  USING SAME
;    NUMBER OF SEQUENCES:  227
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/630,915A
;      FILING DATE:  03-APR-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  1101-174
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1734 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-630-915A-23


  Query Match             3.0%;  Score 32.8;  DB 4;  Length 1734;
  Best Local Similarity   76.9%;  Pred. No. 3.1;
  Matches   40;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy   1049 AATATAAAGATTGATGAACATTATTCACAAAAAAAAAAAAAAAAAGGGCGGC 1100
          || | ||| | |||| || |  |   | ||||||||||||||||||||||||
Db   1680 AACAAAAAAAATGATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGGCGGC 1731


RESULT  39
US-08-306-691B-23
; Sequence 23, Application US/08306691B
; Patent No. 5734039
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Calabretta, Bruno
;    APPLICANT:  Skorski, Tomasz
;    TITLE OF INVENTION:  ANTISENSE
;    TITLE OF INVENTION:  OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
;    NUMBER OF SEQUENCES:  55
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seidel, Gonda, Lavorgna & Monaco, P.C.
;      STREET:  Two Penn Center, Suite 1800
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  U.S.A.
;      ZIP:  19102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 720 Kb
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/306,691B
;      FILING DATE:  September 15, 1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monaco, Daniel A.
;      REGISTRATION NUMBER:  30,480
;      REFERENCE/DOCKET NUMBER:  8321-8
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-8383
;      TELEFAX:  (215) 568-5549
;      TELEX:  No. 5734039e
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2301 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-306-691B-23


  Query Match             3.0%;  Score 32.8;  DB 1;  Length 2301;
  Best Local Similarity   56.5%;  Pred. No. 3.7;
  Matches   61;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy    114 AGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGG 173
          || | || ||  | ||| | |  | |||| ||   |   ||||| ||||||| |||||
Db     27 AGGAGCAGGAGCAGGAGGAGCAGGAGGAGCAGGAGGAGCAGGAGGAGCAGGAGGAGCAGG 86

Qy    174 AGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGG 221
          |  | || |       |   |||| |  ||  ||| | |||   |||
Db     87 AACAGGAGGAGGAGGAGGAGGAGAAGGAGGAGCAGGAAGAGCAGGAGG 134


RESULT  40
PCT-US93-06251-78
; Sequence 78, Application PC/TUS9306251
;  GENERAL INFORMATION:
;    APPLICANT:  Wickstrom, Eric and Rife, Jason P.
;    TITLE OF INVENTION:  Trivalent Synthesis of Oligonucleotides Containing
;    TITLE OF INVENTION:  Stereospecific Alkylphosphonates and Arylphosphonates
;    NUMBER OF SEQUENCES:  93
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCULLY, SCOTT, MURPHY & PRESSER
;      STREET:  400 Garden City Plaza
;      CITY:  Garden City
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  11530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/06251
;      FILING DATE:  19930630
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DiGiglio, Frank S.
;      REGISTRATION NUMBER:  31,346
;      REFERENCE/DOCKET NUMBER:  8586
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  516-742-4343
;      TELEFAX:  516-742-4366
;      TELEX:  230 901 SANS UR
;  INFORMATION FOR SEQ ID NO:  78:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2301 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US93-06251-78


  Query Match             3.0%;  Score 32.8;  DB 5;  Length 2301;
  Best Local Similarity   56.5%;  Pred. No. 3.7;
  Matches   61;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy    114 AGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGG 173
          || | || ||  | ||| | |  | |||| ||   |   ||||| ||||||| |||||
Db     27 AGGAGCAGGAGCAGGAGGAGCAGGAGGAGCAGGAGGAGCAGGAGGAGCAGGAGGAGCAGG 86

Qy    174 AGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGG 221
          |  | || |       |   |||| |  ||  ||| | |||   |||
Db     87 AACAGGAGGAGGAGGAGGAGGAGAAGGAGGAGCAGGAAGAGCAGGAGG 134


Search completed: September 12, 2002, 16:08:14
Job time: 19538 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2002, 09:56:41 ; Search time 2588.38 Seconds
(without alignments)
5751.529 Million cell updates/sec

Title: US-09-824-178-1
Perfect score: 1103
Sequence: 1 GTGACAGGAGCCGAAGCAGC.....AAAAAAAAAAGGGCGGCGCG 1103

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1064.6	96.5	2520	11	AF067170	AF067170 Homo sapi
	2	735.6	66.7	791	10	BG202037	BG202037 RST21384
	3	714.4	64.8	788	10	BI463649	BI463649 603207313
	4	692.4	62.8	862	10	BM048353	BM048353 603625583
	5	686	62.2	898	10	BG330921	BG330921 602431568
	6	686	62.2	939	10	BF684128	BF684128 602141483
	7	652.6	59.2	839	10	BI914640	BI914640 603179407
	8	641.2	58.1	773	10	BI770278	BI770278 603055970
	9	637.4	57.8	729	10	BI670461	BI670461 603294644
	10	634.8	57.6	783	10	BE747764	BE747764 601578767
	11	633.2	57.4	1213	10	BM464928	BM464928 AGENCOURT
	12	625.4	56.7	736	10	BE783278	BE783278 601470834
	13	624.2	56.6	942	10	BF982510	BF982510 602308101
	14	598.2	54.2	612	10	BF997045	BF997045 QV3-GN020
c	15	598	54.2	628	10	BF997056	BF997056 QV3-GN020
c	16	591.2	53.6	611	10	BI015179	BI015179 MR4-ET014
c	17	588.4	53.3	613	10	BF997050	BF997050 QV3-GN020
	18	588	53.3	726	10	BG619227	BG619227 602616238
	19	583.4	52.9	604	10	BF997048	BF997048 QV3-GN020
c	20	575.4	52.2	614	10	BI014315	BI014315 MR4-ET014
c	21	571.6	51.8	600	10	BF997043	BF997043 QV3-GN020
	22	565.6	51.3	591	10	BF997044	BF997044 QV3-GN020
	23	564.6	51.2	638	10	BF852437	BF852437 MR3-EN008
c	24	561.6	50.9	634	10	BF852557	BF852557 MR3-EN008
c	25	560.2	50.8	588	10	BF768839	BF768839 PM0-IT001
	26	557.6	50.6	1070	11	AK011334	AK011334 Mus muscu
	27	556	50.4	1067	11	AK019170	AK019170 Mus muscu
	28	555	50.3	612	10	BF852494	BF852494 MR3-EN008
	29	551.6	50.0	689	10	BM007320	BM007320 603616529
	30	549.6	49.8	2256	11	BC020533	BC020533 Mus muscu
c	31	544.2	49.3	565	10	BF853400	BF853400 MR3-EN008
	32	540.8	49.0	568	10	BM128047	BM128047 if09a04.y
	33	539.4	48.9	563	10	BF768834	BF768834 PM0-IT001
c	34	538.4	48.8	601	10	BF852560	BF852560 MR3-EN008
c	35	535.2	48.5	616	10	BF773208	BF773208 PM0-IT001
c	36	535	48.5	549	10	BF996987	BF996987 QV3-GN020
c	37	531.8	48.2	543	10	BF996978	BF996978 QV3-GN020
c	38	531.8	48.2	576	10	BF997034	BF997034 QV3-GN020
	39	531.4	48.2	725	10	BI258970	BI258970 602971620
	40	524.4	47.5	578	10	BF852439	BF852439 MR3-EN008
	41	523.8	47.5	811	10	BI856660	BI856660 603385820
	42	521.2	47.3	1109	10	BE872419	BE872419 601448686
	43	518	47.0	543	10	BF996983	BF996983 QV3-GN020
	44	514.2	46.6	538	10	BF996982	BF996982 QV3-GN020
	45	511.6	46.4	989	10	BF305079	BF305079 601888405

ALIGNMENTS

RESULT 1
AF067170
LOCUS AF067170 2520 bp mRNA linear HTC 22-MAY-2001
DEFINITION Homo sapiens alpha endosulfine mRNA, complete cds.
ACCESSION AF067170
VERSION AF067170.1 GI:4894373
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2520)
AUTHORS Zhang,Q.H., Ye,M., Wu,X.Y., Ren,S.X., Zhao,M., Zhao,C.J., Fu,G., Shen,Y., Fan,H.Y., Lu,G., Zhong,M., Xu,X.R., Han,Z.G., Zhang,J.W., Tao,J., Huang,Q.H., Zhou,J., Hu,G.X., Gu,J., Chen,S.J. and Chen,Z.
TITLE Cloning and functional analysis of cDNAs with open reading frames for 300 previously undefined genes expressed in CD34+ hematopoietic stem/progenitor cells
JOURNAL Genome Res. 10 (10), 1546-1560 (2000) X bad date
MEDLINE 20499367
PUBMED 11042152
REFERENCE 2 (bases 1 to 650)
AUTHORS Zhang,Q., Fu,G., Wu,J., Zhou,J., Ye,M., Shen,Y., Kan,L., He,K., Gu,B., Chen,S., Mao,M. and Chen,Z.
TITLE Human alpha endosulfine gene
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 2520)
AUTHORS Zhang,Q.
TITLE Direct Submission
JOURNAL Submitted (18-MAY-1998) Shanghai Institute of Hematology, Shanghai Second Medical University, Rui-Jin Hospital, 197 Rui-Jin Road II, Shanghai 200025, China
FEATURES Location/Qualifiers
source 1. .2520
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="CD34+ cell"
CDS 110. .463
/codon_start=1
/product="alpha endosulfine"

```
                     /protein_id="AAD32454.1"
                     /db_xref="GI:4894374"
                     /translation="MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAG"
BASE COUNT      726 a    535 c    611 g    648 t
ORIGIN
```

Query Match 96.5%; Score 1064.6; DB 11; Length 2520;
Best Local Similarity 99.5%; Pred. No. 4.9e-198;
Matches 1078; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

```
Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
Db       4 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 63

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
Db      64 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 123

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
Db     124 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 183

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
Db     184 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 243

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
Db     244 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 303

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
Db     304 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 363

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
Db     364 AGGACCAGACAAGAACCTGGTGACCGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 423

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCC 480
Db     424 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCC 483

Qy     481 CCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCA 540
Db     484 CCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCA 543

Qy     541 GATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGT 600
Db     544 GATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGT 603

Qy     601 AATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACA 660
Db     604 AATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACA 663

Qy     661 CTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAA 720
Db     664 CTCCCTGTAGTCCAGGTAGGGCCCAGAAATGGGAAAGGCTAGGATTGGATAATGCTGCAA 723

Qy     721 ATGC-TTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTC 779
Db     724 ATGCTTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTC 783

Qy     780 CCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAA 839
Db     784 CCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAA 843

Qy     840 CCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTC 899
Db     844 CCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTC 903

Qy     900 TGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAA 959
Db     904 TGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAA 963

Qy     960 AAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTA 1019
Db     964 AAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTA 1023

Qy    1020 TCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACATTATTCACAAA 1079
Db    1024 TCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACATTATTCACAGA 1083

Qy    1080 AAA 1082
Db    1084 AAA 1086
```

RESULT 2
BG202037

```
LOCUS       BG202037                791 bp    mRNA    linear   EST 21-APR-2001
DEFINITION  RST21384 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG202037
VERSION     BG202037.1  GI:13723724
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 791)
  AUTHORS   Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.
            , Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith
            ,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
            ,J., Danzig,J. and Ducar,M.
  TITLE     Creation of genome-wide protein expression libraries using random
            activation of gene expression
  JOURNAL   Nat. Biotechnol. 19 (5), 440-445 (2001)
  MEDLINE   21227151
COMMENT     Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 552.
FEATURES             Location/Qualifiers
     source          1. .791
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Athersys RAGE Library"
                     /cell_line="HT1080"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."
BASE COUNT      215 a    173 c    181 g    217 t      5 others
ORIGIN
```

Query Match 66.7%; Score 735.6; DB 10; Length 791;
Best Local Similarity 98.0%; Pred. No. 9.4e-134;
Matches 774; Conservative 0; Mismatches 13; Indels 3; Gaps 3;

```
Qy     289 GCAAAAGTACTTTGACTC-AGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGC 347
Db       2 GCCAAAGTACTTTGTCTCAAGGAAACTACAACATGGCCAAAGCCAAGATGAAGAATAAGC 61

Qy     348 AGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCAC 407
Db      62 AGCTGCCAAGTGCAGGACCAGACAAGAACCTGGCGACTGGTGATCACATCCCCACCCCAC 121
```

```
Qy     408 AGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGC 467
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 AGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGC 181

Qy     468 CCCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAAT 527
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 CCCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAAT 241

Qy     528 GGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGG 587
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 GGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGG 301

Qy     588 CTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTG 647
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 CTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTG 361

Qy     648 ACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTG 707
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     362 ACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCCAGAAATAGGAAAGGCTAGGATTG 421

Qy     708 GATAATGCTGCAAATGCTTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTG 767
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 GATAATGCTGCAAATGCTTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTG 481

Qy     768 GGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACA 827
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 GGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACA 541

Qy     828 TCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGC 887
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 TCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGC 601

Qy     888 TTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAG 947
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 TTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAG 661

Qy     948 AGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTC 1007
           ||||||||||||||| |||||||||||||||| |||||||||| ||||||||||||||||
Db     662 AGAATCCTTAAAAAGAANAATTGACTTGCTTGCCT-AAACTACAGTG-CCCCTTAGCCTC 719

Qy    1008 CATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAAC 1067
           |||||||||||||||||||||||||||||||| |||||||     |||||||
Db     720 CATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCANATAATATANNAGATGATGAAT 779

Qy    1068 ATTATTCACA 1077
           ||||||||||
Db     780 ATTATTCACA 789
```

RESULT 3
BI463649
LOCUS BI463649 788 bp mRNA linear EST 21-AUG-2001
DEFINITION 603207313F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5273162 5', mRNA sequence.
ACCESSION BI463649
VERSION BI463649.1 GI:15254305
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 788)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11689 row: i column: 03
High quality sequence stop: 788.

FEATURES Location/Qualifiers
source 1. .788
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5273162"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 215 a 204 c 211 g 158 t
ORIGIN

Query Match 64.8%; Score 714.4; DB 10; Length 788;
Best Local Similarity 96.5%; Pred. No. 1.3e-129;
Matches 741; Conservative 0; Mismatches 26; Indels 1; Gaps 1;

```
Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 76

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 136

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     137 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCTGGAGAAAGA 196

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 256

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 316

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 376

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCAGCCCACAGGATCTGCCCCA 420
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     377 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCAGCCCACAGGATCTGCCCCA 436

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     437 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCC 496

Qy     481 CCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     497 CCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCA 556

Qy     541 GATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     557 GATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGT 616

Qy     601 AATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACA 660
            |||||||||| |||||||||||||||||||||||||||||||      | |  |||
Db     617 -ATGTAGGTTGACAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTTGACACTTCCAACA 675

Qy     661 CTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAA 720
```

```
Db    676 GTCCCTGTAGTCCAGGTAGGGGCTAGAAATAGGAAAGGCTAGGATGTGGATATGCTGCAA 735

Qy    721 ATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGG 768

Db    736 AATGCTTTTAGGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTGGG 783


RESULT   4
BM048353
LOCUS       BM048353                862 bp    mRNA    linear   EST 07-NOV-2001
DEFINITION  603625583F1 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:5452053 5',
            mRNA sequence.
ACCESSION   BM048353
VERSION     BM048353.1  GI:16777620
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 862)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1943  row: f  column: 22
            High quality sequence stop: 750.
FEATURES             Location/Qualifiers
     source          1. .862
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5452053"
                     /clone_lib="NIH_MGC_40"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed  by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California,  Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life  Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      238 a    190 c    207 g    227 t
ORIGIN


  Query Match           62.8%;  Score 692.4;  DB 10;  Length 862;
  Best Local Similarity 96.0%;  Pred. No. 2.5e-125;
  Matches  774;  Conservative    0;  Mismatches   26;  Indels    6;  Gaps    6;

Qy    280 CCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAA 339

Db      2 CCAGAAAGGGCAAAAGTAC-TTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAA 60

Qy    340 GAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCC 399

Db     61 GAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCC 120

Qy    400 CACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTA 459

Db    121 CACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTA 180

Qy    460 ACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGG 519

Db    181 ACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGG 240

Qy    520 GATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCT 579

Db    241 GATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCT 300

Qy    580 GCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCG 639

Db    301 GCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCG 360

Qy    640 TTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGC 699

Db    361 TTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGC 420

Qy    700 TAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTC 759

Db    421 TAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTC 480

Qy    760 TCCTTTTGGGAGAGAATGTCCCAA-AATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCA 818

Db    481 TCCTTTTGGGAGAGAATGTCCCAAGAATTGATTAGGCTGAGCCTTGGGAATAGTCTGGCA 540

Qy    819 GGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGT 878

Db    541 GGTTTAACATCCCAAGGCTAA-CTAACGTAGTTGGGACAGGTAGATTGAATGAGACATGT 599

Qy    879 TTTCTGTGCTTCTAAGTGTTCTGT-CCCTTAGGCTGCTATTGCTTCATGTTTCCATTATG 937

Db    600 TTTCTGTGCTTCTAAGTGTTCTGTACCCTTAGGCTGCTATCGCTTCATGTGTCCATTATG 659

Qy    938 GCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCC 997

Db    660 GCAGGCTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGGCTAAAACTACAGTGCCC 719

Qy    998 CCTTAGCCTCCATTACTTA-GTATCTCTTACAGTTTGCT-CTGGCTCTCAAATAATATAA 1055

Db    720 CCTTAAGCTCCATTACTTAGGTATCTCTTACAGATGGCTCCGGGGTCTCAAATCACTATA 779

Qy   1056 AGATTGATGAACATTATTCACAAAAA 1081

Db    780 ACAGATGATGACATTATTCACGAAAA 805


RESULT   5
BG330921
LOCUS       BG330921                898 bp    mRNA    linear   EST 27-FEB-2001
DEFINITION  602431568F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4549289 5',
            mRNA sequence.
ACCESSION   BG330921
VERSION     BG330921.1  GI:13137359
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 898)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1240  row: g  column: 18
            High quality sequence stop: 662.
FEATURES             Location/Qualifiers
     source          1. .898
                     /organism="Homo sapiens"
```

/db_xref="taxon:9606"
/clone="IMAGE:4549289"
/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 236 a 254 c 234 g 174 t
ORIGIN

Query Match 62.2%; Score 686; DB 10; Length 898;
Best Local Similarity 96.7%; Pred. No. 4.4e-124;
Matches 733; Conservative 0; Mismatches 20; Indels 5; Gaps 3;

```
Qy      10 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 GCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCT 61

Qy      70 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 121

Qy     130 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 181

Qy     190 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 241

Qy     250 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 301

Qy     310 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 369
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 361

Qy     370 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 429
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 421

Qy     430 CTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCA 489
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 CTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCA 481

Qy     490 ACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTC 549
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 ACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTC 541

Qy     550 AGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGT 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 AGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGT 601

Qy     610 TGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCTGTA 669
           || |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     602 TGCAAAGGGGCTCTAAGTCTATTTATTTCGTTAGATTGACACTTCCACACACTCCCTGTA 661

Qy     670 GTCCAGG---TAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGCTT 726
           || || |     || ||   |||||||| |||||||| ||||||||||||||||||| |
Db     662 GTTCAAGGTAGGGGCCCAGCAAATAGGACAGGCTAGGA-TGGATAATGCTGCAAATGC-T 719

Qy     727 TTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTT 764
           |||||| |||||||   |     ||||||||||||||
Db     720 TTTTTGCGTGAGAACTGGAGACAATGTGATTTCTCCTT 757
```

RESULT 6
BF684128
LOCUS BF684128 939 bp mRNA linear EST 22-DEC-2000
DEFINITION 602141483F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4302356 5', mRNA sequence.
ACCESSION BF684128
VERSION BF684128.1 GI:11969536
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 939)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1164 row: f column: 21
High quality sequence stop: 697.

FEATURES Location/Qualifiers
source 1. .939
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4302356"
/clone_lib="NIH_MGC_46"
/tissue_type="leiomyosarcoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 245 a 260 c 250 g 184 t
ORIGIN

Query Match 62.2%; Score 686; DB 10; Length 939;
Best Local Similarity 91.5%; Pred. No. 4.3e-124;
Matches 761; Conservative 0; Mismatches 65; Indels 6; Gaps 3;

```
Qy       5 CAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGT 64
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 CAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGT 61

Qy      65 TGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAA 124
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 TGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAA 121

Qy     125 GAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGT 184
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 GAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGT 181

Qy     185 ATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAA 244
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 ATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAA 241

Qy     245 AAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGAC 304
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 AAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGAC 301
```

```
Qy    305 TCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGA 364
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 TCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGA 361

Qy    365 CCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGA 424
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    362 CCAGACAAGAACCTGGTGACTGGTGATCACATCCCCAGCCCACAGGATCTGCCCCAGAGA 421

Qy    425 AAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTT 484
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 AAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTT 481

Qy    485 CCTCAACCACTGGACGTTTATATATTATA--GGCAGGGATGAAATGGGCACCTAGTCAGA 542
          |||||||||||||||||||||||||||||  |   || |||| ||||||||||||||||||
Db    482 CCTCAACCACTGGACGTTTATATATTATACGGCAGGGCATGACATGGGCACCTAGTCAGA 541

Qy    543 TCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAA 602
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 TCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAA 601

Qy    603 TGTAGGTTGAAA--AGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACA 660
          ||||||  |  ||  |||||||||||||||||| || |||||||||||||||||||||||
Db    602 TGTAGGGTTGAACCAGGGGCTCTAAGTTTATTCATATCGTTAGATTGACACTTCCACACA 661

Qy    661 CTCCCTGTAGTCCAGGTAGGGCCTAGAAA--TAGGAAAGGCTAGGATTGGATAATGCTGC 718
          |||||||||||||||||||||||| |||||    | || |||||||| || | |||||||
Db    662 CTCCCTGTAGTCCAGGTAGGGCCCAGAAACTAGGAAAGGGCTAGGATCGGCTCATGCTGC 721

Qy    719 AAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGT 778
          |||||   ||| ||||     ||||||          ||| |||||||| |  ||||
Db    722 AAATGGCATTTCTGTGCTGAGAACTGGGAGAGATGTGATTCCCCTTTTGGGCGCAAATGG 781

Qy    779 CCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCC 830
          |||||  |  | ||||||    ||| || |  |  ||  |  |  ||||  |    ||
Db    782 TCCAAAGTGGCTTAGGCGAGCCCTGGGCACTCGTGGGACGGGTTACATTCCC 833
```

```
RESULT   7
BI914640
LOCUS       BI914640                 839 bp    mRNA    linear   EST 16-OCT-2001
DEFINITION  603179407F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5243713 5',
            mRNA sequence.
ACCESSION   BI914640
VERSION     BI914640.1  GI:16198993
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 839)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11614  row: n  column: 02
            High quality sequence stop: 710.
FEATURES             Location/Qualifiers
     source          1. .839
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5243713"
                     /clone_lib="NIH_MGC_121"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
                     fetal brains, female age 20 weeks, female age 24 weeks,
                     and male age 26 weeks.  Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed  upon
                     cloning).  Average insert size 1.7 kb, insert size range
                     0.7-3.5 kb.  Library is normalized and enriched for
                     full-length clones and  was constructed by C. Gruber
                     (Invitrogen).  Research Genetics  tracking code 017. Note:
                     this is a NIH_MGC Library."
BASE COUNT      236 a    218 c    221 g    164 t
ORIGIN
```

```
  Query Match             59.2%;  Score 652.6;  DB 10;  Length 839;
  Best Local Similarity   92.9%;  Pred. No. 1.5e-117;
  Matches  774;  Conservative    0;  Mismatches   44;  Indels   15;  Gaps    8;

Qy      9 AGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCC 68
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCC 60

Qy     69 TTCCCGGGCCCCTTACACTCCACAG-TCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAA 127
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     61 TTCCCGGGCCCCTTACACTCCACAGATCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAA 120

Qy    128 GAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATT 187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATT 180

Qy    188 CTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAG 247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAG 240

Qy    248 CCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCA 307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCA 300

Qy    308 GGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCA 367
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCA 360

Qy    368 GACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAG 427
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    361 GACAAGAACCTGGTGACTGGTGATCACATCCCCAGCCCACAGGATCTGCCCCAGAGAAAG 420

Qy    428 TCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCT 487
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    421 TCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGACCCCCTCTCCTCCCCTTCCT 480

Qy    488 CAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTC 547
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTC 540

Qy    548 TCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGT--AATGT 605
          |||||||||||||||||||||||||||||||||||||||||||||||||||  | |||
Db    541 TCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTACACTGT 600

Qy    606 AGGTTG--AAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTT-CCACACACT 662
          ||||||  | ||||||||||||||||||| ||||||||||||||||||| |||||||||
Db    601 AGGTTGCACACAGGGGCTCTAAGTTTATTCATTTCGTTAGATTGACACTTCCCACACACT 660

Qy    663 CCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGG--ATAATGCTGCAA 720
          |||||||||||||||||||| |||||||||||||||||||||||    |    |||||
Db    661 CCCTGTAGTCCAGGTAGGGCC-AGAAATAGGAAAGGCTAGGATTGGCATAATGCTTGCAA 719

Qy    721 ATGCTTTTTTTGTGTGAGAAA----CTGGAGAGATGTGATTTCTCCTTTTGGGAGA--GA 774
          |||| ||    | ||||| ||       ||||      || ||||||| |||||||   |
Db    720 ATGCGTTAGCGGGGTGAGGAAAATTGGAGAGACTGTGAATGTCTCCTTATGGGAGAAGAA 779

Qy    775 ATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACA 827
          ||||  |  |||||||||||||||||| ||||||||||  ||| |  | | |
Db    780 TGGTCCAAGCATTGATTAGGCTGAGCCTGGGGAATAGTTGAGCAAGGATTAAA 832
```

```
RESULT   8
BI770278
LOCUS       BI770278                773 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603055970F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5205640 5',
            mRNA sequence.
ACCESSION   BI770278
VERSION     BI770278.1  GI:15761856
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 773)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11515  row: k  column: 17
            High quality sequence stop: 712.
FEATURES             Location/Qualifiers
     source          1. .773
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5205640"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20-22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon  cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was  constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking  code 026. Note:
                     this is a NIH_MGC Library."
BASE COUNT      208 a    227 c    193 g    145 t
ORIGIN


  Query Match           58.1%;  Score 641.2;  DB 10;  Length 773;
  Best Local Similarity   94.7%;  Pred. No. 2.5e-115;
  Matches  708;  Conservative    0;  Mismatches   33;  Indels    7;  Gaps    4;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 85

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 145

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     146 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 205

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     206 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 265

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     266 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 325

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     326 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 385

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     386 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 445

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     446 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCC 505

Qy     481 CCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCA 540
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     506 CCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGACATGGGCACCTAGTCA 565

Qy     541 GATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     566 GATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGT 625

Qy     601 AATGTAGGTTGAAAAGGGG---CTCTAAGTTTATTTATTTCGTTAGATTGACACTTCC-A 656
           |||||||||||| |  |||   ||  |  |   |  | ||||||||||||||||||| |
Db     626 CATGTAGGTTGACACAGGGGCTCTACAGTTCATTCATTCTCGTTAGATTGACACTTCCAA 685

Qy     657 CACACTCCCTGTAG-TCCAGGTAGGGCCTAGAAATAGGAAA--GGCTAGGATTGGATAAT 713
           |||||||||||||| |||||||||||||||||||| || |  |     ||| |||||||
Db     686 CACACTCCCTGTAGATCCAGGTAGGGCCTAGAAATACGACACGGCTAGCGATGGGATAAT 745

Qy     714 GCTGCAAATGCTTTTTTTGTGTGAGAAA 741
           |||||| |||| ||    | |||| | |
Db     746 GCTGCACATGCCTTCACCGCGTGACACA 773


RESULT   9
BI670461
LOCUS       BI670461                729 bp    mRNA    linear   EST 12-SEP-2001
DEFINITION  603294644F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5313851 5',
            mRNA sequence.
ACCESSION   BI670461
VERSION     BI670461.1  GI:15584694
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 729)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11795  row: h  column: 12
            High quality sequence stop: 711.
FEATURES             Location/Qualifiers
     source          1. .729
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5313851"
                     /clone_lib="NIH_MGC_96"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.3 kb and
                     normalized to ROT 5. This is a primary library enriched
```

for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 196 a 201 c 193 g 139 t
ORIGIN

Query Match 57.8%; Score 637.4; DB 10; Length 729;
Best Local Similarity 96.7%; Pred. No. 1.4e-114;
Matches 672; Conservative 0; Mismatches 21; Indels 2; Gaps 2;

```
Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           |||||||||| |||||||||||||||| |||||||||||||||||||||||||||||||
Db      34 GTGACAGGAGTCGAAGCAGCAGCGCAGG-TGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 92

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAAC- 119
           |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||| |
Db      93 CGGTTGCCTTCCAGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAGCG 152

Qy     120 AAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAG 179
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     153 GGGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAG 212

Qy     180 AAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAG 239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     213 AAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAG 272

Qy     240 GACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACT 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     273 GACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACT 332

Qy     300 TTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTG 359
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     333 TTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTG 392

Qy     360 CAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCC 419
           |||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||
Db     393 CAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCATCCCAGAGGATCTGCCCC 452

Qy     420 AGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTC 479
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     453 AGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTC 512

Qy     480 CCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTC 539
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     513 CCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTC 572

Qy     540 AGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGG 599
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     573 AGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGG 632

Qy     600 TAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACAC 659
           |||||||||||||  |  |||     |  ||  |||||||||||||||||||||||||||
Db     633 TAATGTAGGTTGACACAGGGGCTCTAAGTTTATTATTTCGTTAGATTGACACTTCCACAC 692

Qy     660 ACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGA 694
           |||||||||||||||||| || ||||||||||||
Db     693 ACTCCCTGTAGTCCAGGTAAGGGCTAGAAATAGGA 727
```

RESULT 10
BE747764
LOCUS BE747764 783 bp mRNA linear EST 15-SEP-2000
DEFINITION 601578767F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3927545 5', mRNA sequence.
ACCESSION BE747764
VERSION BE747764.1 GI:10161756
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 783)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM757 row: e column: 18
High quality sequence start: 24
High quality sequence stop: 767.

FEATURES Location/Qualifiers
source 1. .783
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3927545"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 220 a 180 c 218 g 165 t
ORIGIN

Query Match 57.6%; Score 634.8; DB 10; Length 783;
Best Local Similarity 96.0%; Pred. No. 4.4e-114;
Matches 716; Conservative 0; Mismatches 22; Indels 8; Gaps 6;

```
Qy      70 TCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 129
           | ||  | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38 TTCCCTGGCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGA 97

Qy     130 GAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 189
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      98 GAACCCTGCGGAGGAGACCGGTGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCT 157

Qy     190 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     158 GCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCC 217

Qy     250 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     218 TGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGG 277

Qy     310 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 369
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     278 AGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA 337

Qy     370 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 429
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     338 CAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTC 397

Qy     430 CTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCA 489
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     398 CTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCA 457

Qy     490 ACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTC 549
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     458 ACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTC 517

Qy     550 AGCTTGCTAGCCAGAAATGACTGTGATTCTGCT-GGGGGCTGCTGAGAAGGTAATGTAGG 608
           |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
```

```
Db    518 AGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGGCTGCTGAGAAGGTAATGTAGG 577

Qy    609 TTGAAAA-GGGGCTCTAAGTTTATTTATTTC-GTTAGATTGACACTTCCACACACTCCCT 666
          ||||||| |||||||||||||||||||||| |||||||||||||||||| |||||||||
Db    578 TTGAAAAGGGGGCTCTAAGTTTATTTATTTCGGTTAGATTGACACTTCCAAACACTCCCT 637

Qy    667 GTAGTCCAGGTAGGGCCTAGAAATA--GGAAAGGCTAGGATTGGATAATGCTGCAAATGC 724
          ||||||||||||||| ||||||||| | || |||||||||||||||||| || ||||||
Db    638 GTAGTCCAGGTAGGGTCTAGAAATACGGAAAGGGCTAGGATTGGATAATGTTGAAAATGC 697

Qy    725 TTTTTTTGTG--TGAGAAACTGGAGAGATGTGA-TTTCTCCTTTTGGGAGAGAATGTCCC 781
          ||||||||||     ||  |||||||||||| ||||||||||||||||||||||||||
Db    698 TTTTTTTGTGGTGAGAAACTTGGAGAGATGTGATTTTCTCCTTTTGGGAGAGAATGTCCC 757

Qy    782 AAAATTGATTAGGCTGAGCCTTGGGA 807
          |||||||||||||     |||||||
Db    758 AAAATTGATTAGGGTTGAGCTTGGGA 783


RESULT  11
BM464928
LOCUS       BM464928                1213 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6428376 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5505048
            5', mRNA sequence.
ACCESSION   BM464928
VERSION     BM464928.1  GI:18513970
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1213)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lou Staudt
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12147  row: g  column: 01
            High quality sequence start: 3
            High quality sequence stop: 577.
FEATURES             Location/Qualifiers
     source          1. .1213
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5505048"
                     /clone_lib="NIH_MGC_85"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.867 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      333 a    295 c    301 g    284 t
ORIGIN


  Query Match             57.4%;  Score 633.2;  DB 10;  Length 1213;
  Best Local Similarity   97.2%;  Pred. No. 8.3e-114;
  Matches  655;  Conservative    0;  Mismatches   18;  Indels    1;  Gaps    1;

Qy    187 TCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAA 246
          || ||    | | |  ||||||||||||||||||||||||||||||||||||||||||||
Db     26 TCCGCGGACGCGTGGGGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAA 85

Qy    247 GCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTC 306
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     86 GCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTC 145

Qy    307 AGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACC 366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    146 AGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACC 205

Qy    367 AGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAA 426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    206 AGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAA 265

Qy    427 GTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCC 486
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    266 GTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCC 325

Qy    487 TCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTT 546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    326 TCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTT 385

Qy    547 CTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTA 606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    386 CTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTA 445

Qy    607 GGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCT 666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    446 GGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCT 505

Qy    667 GTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGCTT 726
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    506 GTAGTCCAGGTAGGGCCCAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGCTT 565

Qy    727 TTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCCAAAAT 786
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    566 TTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCCAAAT 625

Qy    787 TGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACAT-CCCAAGGCTAACCTAAC 845
          |||||||| |||||||||||||| |||||||||||||||||||| |||| |||||||| |||
Db    626 TGATTAGGGTGAGCCTTGGGAAAAGTTTGGCAGGTTTAACATCCCCAGGGCTAACCCAAC 685

Qy    846 GTAGTTGGGAAAGG 859
          | ||||||| || |
Db    686 GGAGTTGGGGAAAG 699


RESULT  12
BE783278
LOCUS       BE783278                736 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601470834F1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3873753 5',
            mRNA sequence.
ACCESSION   BE783278
VERSION     BE783278.1  GI:10204476
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 736)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9630  row: d  column: 10
            High quality sequence stop: 669.
FEATURES             Location/Qualifiers
     source          1. .736
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3873753"
                     /clone_lib="NIH_MGC_67"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      204 a    185 c    192 g    155 t
ORIGIN

  Query Match           56.7%;  Score 625.4;  DB 10;  Length 736;
  Best Local Similarity  96.9%;  Pred. No. 3e-112;
  Matches  712;  Conservative    0;  Mismatches   16;  Indels    7;  Gaps    7;

Qy      39 TTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCG 98
           || |||  |  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTTCCCCTCGGCTTCCCTTCTCCGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCG 60

Qy      99 GTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGA 158
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GTCCCGCCATGTCCCAGAAACAAGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGA 120

Qy     159 AGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAA 218
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAA 180

Qy     219 AGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGAC 278
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGAC 240

Qy     279 TCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGA 338
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     241 TCCAGAAAGGGCAAAAGTAC-TTGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGA 299

Qy     339 AGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCC 398
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 AGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCACATCC 359

Qy     399 CCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGT 458
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 CCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGT 419

Qy     459 AACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAG 518
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 AACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAG 479

Qy     519 GGATGAAATGGGCACCTAGTCAGATCTTCTCAGCT-TGCTAGCCAGAAATGACTGTGATT 577
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     480 GGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTGTGCTAGCCAGAAATGACTGTGATT 539

Qy     578 CTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAG-TTTATTTATT 636
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     540 CTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTTATTTATT 599

Qy     637 TCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAA 696
           |||||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||
Db     600 TCGTTAGATTGACACTTCCA-ACACTCCCTGTAGTCCAGGTAGGG-CTAGAAATAGGAAA 657

Qy     697 GGCTAGGATTGGATAATGCTGCAAATGCTTTTTTTTGTGTGAGAAACTGGAGAGATGTGAT 756
           ||||| || ||||| ||||||  |||||||||||||||||||||||||||||||||||||||
Db     658 GGCTACGA-TGGATTATGCTGGAAATGC-TTTTTTGTGTGAGAAACTGGAGAGATGTGAA 715

Qy     757 TTCTCCTTTTGGGAG 771
           ||    ||  | |||
Db     716 TTTCTTTTGGGAGAG 730


RESULT  13
BF982510
```

```
LOCUS       BF982510                942 bp    mRNA    linear   EST 23-JAN-2001
DEFINITION  602308101F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4399368 5',
            mRNA sequence.
ACCESSION   BF982510
VERSION     BF982510.1  GI:12385322
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 942)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10103  row: a  column: 01
            High quality sequence stop: 630.
FEATURES             Location/Qualifiers
     source          1. .942
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4399368"
                     /clone_lib="NIH_MGC_88"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      310 a    225 c    212 g    195 t
ORIGIN

  Query Match           56.6%;  Score 624.2;  DB 10;  Length 942;
  Best Local Similarity  85.4%;  Pred. No. 5e-112;
  Matches  781;  Conservative    0;  Mismatches  118;  Indels   16;  Gaps    7;

Qy     155 GAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTG-AGAGAGCTGAAGAGGCAAA 213
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1 GAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGCCTGTAGAGAGCTGAAGAGGCAAA 60

Qy     214 GCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAA 273
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTGGAGGCTCCGACTTCCTCATGAA 120

Qy     274 GAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAA 333
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAGACTACAACATGGCCAAAGCCAA 180

Qy     334 GATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCA 393
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACAAGAACCTGGTGACTGGTGATCA 240

Qy     394 CATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGC 453
           |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CATCCCCAACCCACAGGATCTGCCCCAGAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGC 300

Qy     454 GGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATA 513
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATA 360

Qy     514 GGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGT 573
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGT 420
```

```
Qy     574 GATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTT 633
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTT 480

Qy     634 ATTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGG 693
           |||||||||||||||||||||||||||||||||||||||||||||||||||| | |||||||||
Db     481 ATTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGACCAGAAATAGG 540

Qy     694 AAAGGCTAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGG-AGAGATG 752
           |||||||||||| ||||||||||||||||||      |||||||||||||| |||||||
Db     541 AAAGGCTAGGATCGGATAATGCTGCAAATGCTAACACAGTGTGAGAAACTGGCAGAGATG 600

Qy     753 TGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGT 812
           ||||||||||   |||||||||||| ||||||| ||||||||||||||||||| | ||||
Db     601 TGATTTCTCCTAACGGGAGAGAATGT-CCAAAATCGATTAGGCTGAGCCTTGGCACTAGT 659

Qy     813 TTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAG 872
           |||   |  |      ||||||||||| |    | | || |  |  |     | ||||
Db     660 ATGGACAGGCTCTACATCCAAGGCTAACTAACGTAAATGGGAACGGTAAATGGAACTGAG 719

Qy     873 ACATGTTTTCTGTGCT-TCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCAT-----G 926
           ||||| || ||||||| ||||||   ||| |||        |  |   | |||     |
Db     720 ACATGATTACTGTGCTACCTAAGTAGACTGACCCCCTTAAGCCAGTATATGCATACAACG 779

Qy     927 TTTCCATTATGGCAGGTTTAGAGAA-----TCCTTAAAAAGAAAAATTGACTTGCTTGC- 980
           || |||| |  |      || | ||     |    ||||   ||| ||||||| || ||
Db     780 TTCCCATAAATGGCAAGATATACAAAATTCTTTCAAAAAGACAAATTTGACTTCCTAGCG 839

Qy     981 -CTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGG 1039
           ||| ||| ||   |||||| ||| ||||| |||| |    |  |||||| | |    |
Db     840 ACTAGAACAACGTAGCCCCCATAGACTCCAATACTAAAATACCATTACAGATAGTCTAGC 899

Qy    1040 CTCTCAAATAATATA 1054
           |   |||||| ||
Db     900 ACCCACAATAATCTA 914
```

RESULT 14
BF997045
LOCUS BF997045 612 bp mRNA linear EST 23-JAN-2001
DEFINITION QV3-GN0201-081100-437-c06 GN0201 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF997045
VERSION BF997045.1 GI:12403368
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 612)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV3&t2=QV3-GN0201-081100-437-c06&t3=2000-11-08&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 517.
FEATURES Location/Qualifiers
source 1. .612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0201"
/dev_stage="Adult"
/note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI ; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 159 a 126 c 143 g 184 t
ORIGIN

Query Match 54.2%; Score 598.2; DB 10; Length 612;
Best Local Similarity 98.7%; Pred. No. 6.4e-107;
Matches 603; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy     442 CAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGT 501
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 CAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGT 61

Qy     502 TTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCC 561
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 TTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCC 121

Qy     562 AGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCT 621
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 AGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCT 181

Qy     622 CTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGG 681
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 CTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGG 241

Qy     682 CCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAA 741
           || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 CCCAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAA 301

Qy     742 CTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCC 801
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 CTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCC 361

Qy     802 TTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTA 861
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 TTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTA 421

Qy     862 GATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCT 921
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     422 GATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATAGCT 481

Qy     922 TCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCC 981
           |||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||||
Db     482 TCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAACAAGACAAATTGACTTGCTTGCC 541

Qy     982 TAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGCT 1041
           |||||||||||||  ||||||||||||||||||||||||||||||||||||| ||||||||||
Db     542 TAAAACTACAGTGACCCCTTAGCCTCCATTACTTAGTATCTCTTACAGTATGCTCTGGCT 601

Qy    1042 CTCAAATAATA 1052
           ||||||  |||
Db     602 CTCAAAATATA 612
```

RESULT 15
BF997056/c
LOCUS BF997056 628 bp mRNA linear EST 23-JAN-2001
DEFINITION QV3-GN0201-081100-437-f11 GN0201 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF997056
VERSION BF997056.1 GI:12403379

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV3&t2=QV3-GN0201-
            081100-437-f11&t3=2000-11-08&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 612.
FEATURES             Location/Qualifiers
     source          1. .628
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="GN0201"
                     /dev_stage="Adult"
                     /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                     ; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
BASE COUNT      189 a    147 c    129 g    163 t
ORIGIN
```

```
  Query Match           54.2%;  Score 598;  DB 10;  Length 628;
  Best Local Similarity   98.2%;  Pred. No. 6.9e-107;
  Matches  615;  Conservative    0;  Mismatches   10;  Indels    1;  Gaps    1;

Qy     445 CAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTA 504
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     628 CAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTA 569

Qy     505 TATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGA 564
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     568 TATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGA 509

Qy     565 AATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTA 624
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     508 AATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTA 449

Qy     625 AGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCT 684
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     448 AGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCC 389

Qy     685 AGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGC-TTTTTTTTGTGTGAGAAACT 743
           |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db     388 AGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGCTTTTTTTTTGTGTGAGAAACT 329

Qy     744 GGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTT 803
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     328 GGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTT 269

Qy     804 GGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGA 863
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     268 GGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGA 209

Qy     864 TTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTC 923
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     208 TTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTC 149

Qy     924 ATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTA 983
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     148 ATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTA 89

Qy     984 AAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCT 1043
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88 AAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCT 29

Qy    1044 CAAATAATATAAAGATTGATGAACAT 1069
           |||||||||   ||   ||  | |||
Db      28 CAAATAATAACGAGTCCGAGAACCAT 3
```

```
RESULT  16
BI015179/c
LOCUS       BI015179                611 bp    mRNA    linear   EST 13-JUN-2001
DEFINITION  MR4-ET0140-190201-006-c07 ET0140 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI015179
VERSION     BI015179.1  GI:14419250
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR4&t2=MR4-ET0140-
            190201-006-c07&t3=2001-02-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 6
            High quality sequence stop: 611.
FEATURES             Location/Qualifiers
     source          1. .611
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="ET0140"
                     /dev_stage="Adult"
                     /note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
```

```
BASE COUNT      186 a    143 c    130 g    152 t
ORIGIN


  Query Match             53.6%;  Score 591.2;  DB 10;  Length 611;
  Best Local Similarity   99.3%;  Pred. No. 1.5e-105;
  Matches  604;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy    441 CCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACG 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    611 CCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACG 552

Qy    501 TTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGC 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    551 TTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGC 492

Qy    561 CAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGC 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    491 CAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGC 432

Qy    621 TCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGG 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    431 TCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGG 372

Qy    681 GCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAA 740
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    371 GCCCAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAA 312

Qy    741 ACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGC 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311 ACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGC 252

Qy    801 CTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGT 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    251 CTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGT 192

Qy    861 AGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGC 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    191 AGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGC 132

Qy    921 TTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGC 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    131 TTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGC 72

Qy    981 CTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTATCTCTTA-CAGTTTGCTCTGG 1039
          ||| |||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db     71 CTACAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTATCTCTTACCAGTTTGCTCTGG 12

Qy   1040 CTCTCAAA 1047
          |||||| |
Db     11 CTCTCATA 4


RESULT  17
BF997050/c
LOCUS       BF997050                613 bp    mRNA    linear   EST 23-JAN-2001
DEFINITION  QV3-GN0201-081100-437-e09 GN0201 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF997050
VERSION     BF997050.1  GI:12403373
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV3&t2=QV3-GN0201-
            081100-437-e09&t3=2000-11-08&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 573.
FEATURES             Location/Qualifiers
     source          1. .613
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="GN0201"
                     /dev_stage="Adult"
                     /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                     ; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
BASE COUNT      187 a    139 c    125 g    161 t      1 others
ORIGIN


  Query Match             53.3%;  Score 588.4;  DB 10;  Length 613;
  Best Local Similarity   98.8%;  Pred. No. 5.2e-105;
  Matches  592;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy    467 CCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAA 526
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    604 CCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGNCAGGGATGAAA 545

Qy    527 TGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGG 586
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    544 TGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGG 485

Qy    587 GCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATT 646
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    484 GCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATT 425

Qy    647 GACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATT 706
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    424 GACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCCAGAAATAGGAAAGGCTAGGATT 365

Qy    707 GGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTT 766
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364 GGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTT 305

Qy    767 GGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAAC 826
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304 GGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAAC 245

Qy    827 ATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTG 886
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    244 ATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTG 185

Qy    887 CTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTA 946
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184 CTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTA 125

Qy    947 GAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCT 1006
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    124 GAGAATCCTTAAAAAGAAAAATTGACTTGCCTGCCTAAAACTACAGTGCCCCCTTAGCCT 65
```

```
Qy   1007 CCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGA 1065
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64 CCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAACGATTCGAGA 6
```

```
RESULT  18
BG619227
LOCUS       BG619227                  726 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602616238F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4730240 5',
            mRNA sequence.
ACCESSION   BG619227
VERSION     BG619227.1  GI:13670598
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 726)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLCM1586  row: k  column: 09
           High quality sequence stop: 709.
FEATURES             Location/Qualifiers
     source          1. .726
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4730240"
                     /clone_lib="NIH_MGC_79"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: placenta; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      195 a    211 c    185 g    135 t
ORIGIN
```

```
  Query Match             53.3%;  Score 588;  DB 10;  Length 726;
  Best Local Similarity   96.2%;  Pred. No. 6e-105;
  Matches  666;  Conservative    0;  Mismatches   20;  Indels    6;  Gaps    6;

Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      38 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 97

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      98 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 157

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     158 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 217

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     218 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 277
```

```
Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     278 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 337

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     338 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 397

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     398 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCA-CCCACAGGATCTGCCCCA 456

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCC 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     457 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGT-ACCTGAGCCCCCCTCTCCTCA 515

Qy     481 CCTTCCTCAACCACTGGA-CGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAG-T 538
           ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||| |
Db     516 CCTTCCTCAACCACTGGACCGTTTATATATTATAGGCAGGGATGAAATGGGCCCCTAGTT 575

Qy     539 CAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGC-TGAGAA 597
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     576 CAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTTGAGAC 635

Qy     598 GGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCAC 657
           || ||||||| |  ||| |||     |  ||  ||||||||||||||||||||||||||
Db     636 CGTCATGTAGGGTCGAAAAGGGGCTCTAAGTTCATTATTTCGTTAGATTGACACTTCCA- 694

Qy     658 ACACTCCCTGTAGTCCAGGTAGGGCCTAGAAA 689
           |||||||||||||||| |||||||| |||||
Db     695 ACACTCCCTGTAGTCCACGTAGGGCCCAGAAA 726
```

```
RESULT  19
BF997048
LOCUS       BF997048                  604 bp    mRNA    linear   EST 23-JAN-2001
DEFINITION  QV3-GN0201-081100-437-d11 GN0201 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF997048
VERSION     BF997048.1  GI:12403371
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 604)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV3&t2=QV3-GN0201-
            081100-437-d11&t3=2000-11-08&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 603.
FEATURES             Location/Qualifiers
     source          1. .604
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /clone_lib="GN0201"
                     /dev_stage="Adult"
                     /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                     ; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
BASE COUNT      153 a    123 c    142 g    184 t      2 others
ORIGIN


  Query Match             52.9%;  Score 583.4;  DB 10;  Length 604;
  Best Local Similarity   99.8%;  Pred. No. 4.9e-104;
  Matches  584;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    442 CAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGT 501
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     20 CAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGT 79

Qy    502 TTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCC 561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     80 TTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCC 139

Qy    562 AGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCT 621
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    140 AGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCT 199

Qy    622 CTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGG 681
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 CTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGG 259

Qy    682 CCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAA 741
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    260 CCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAA 319

Qy    742 CTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCC 801
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    320 CTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCC 379

Qy    802 TTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTA 861
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    380 TTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTA 439

Qy    862 GATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCT 921
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    440 GATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCT 499

Qy    922 TCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCC 981
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    500 TCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATCGACTTGCTTGCC 559

Qy    982 TAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTATCTCTTA 1026
          |||||||||||||||||||||||||||||||||||||||||||||
Db    560 TAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTATCTCTTA 604


RESULT  20
BI014315/c
LOCUS       BI014315                614 bp    mRNA    linear   EST 13-JUN-2001
DEFINITION  MR4-ET0140-310101-002-f06 ET0140 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI014315
VERSION     BI014315.1  GI:14418386
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 614)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR4&t2=MR4-ET0140-
            310101-002-f06&t3=2001-01-31&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 614.
FEATURES             Location/Qualifiers
     source          1. .614
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="ET0140"
                     /dev_stage="Adult"
                     /note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      187 a    142 c    126 g    158 t      1 others
ORIGIN


  Query Match             52.2%;  Score 575.4;  DB 10;  Length 614;
  Best Local Similarity   98.5%;  Pred. No. 1.8e-102;
  Matches  601;  Conservative    0;  Mismatches    7;  Indels    2;  Gaps    2;

Qy    453 CGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTAT 512
          |||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    614 CGGGTTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCGACCACTGGACGTTTATATATTAT 555

Qy    513 AGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTG 572
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    554 AGNCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTG 495

Qy    573 TGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATT 632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    494 TGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATT 435

Qy    633 TATTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAG 692
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    434 TATTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCCAGAAATAG 375

Qy    693 GAAAGGCTAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATG 752
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    374 GAAAGGCTAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATG 315

Qy    753 TGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGT 812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    314 TGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGT 255

Qy    813 TTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAG 872
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    254 TTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAG 195

Qy    873 ACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCA 932
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    194 ACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCA 135

Qy    933 TTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAG 992
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    134 TTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAG 75

Qy    993 TGCCCCCTTAGCCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATA 1052
          |||||||||||||||||||||||||||||||| ||||||||||||||| |||||||
Db     74 CGCCCCCTTAGCCTCCATTACTTAGTATCTCTTAC-GTTTGCTCTGGCTCTC-AATAATA 17

Qy   1053 TAAAGATTGA 1062
          ||| ||| ||
Db     16 TAACGATCGA 7


RESULT 21
BF997043/c
LOCUS       BF997043                600 bp    mRNA    linear   EST 23-JAN-2001
DEFINITION  QV3-GN0201-081100-437-b06 GN0201 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF997043
VERSION     BF997043.1  GI:12403366
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV3&t2=QV3-GN0201-
            081100-437-b06&t3=2000-11-08&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 600.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="GN0201"
                     /dev_stage="Adult"
                     /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                     ; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
BASE COUNT      187 a    136 c    121 g    156 t
ORIGIN


  Query Match           51.8%;  Score 571.6;  DB 10;  Length 600;
  Best Local Similarity  98.5%;  Pred. No. 9.9e-102;
  Matches  577;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy    469 CCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATG 528
          |||| |  |||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    600 CCCCCTCTCCTCCCTTCCTCAACCACTGGACGTTTATATATTATAGTCAGGGATGAAATG 541

Qy    529 GGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGC 588
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 GGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGC 481

Qy    589 TGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGA 648
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    480 TGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGA 421

Qy    649 CACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGG 708
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    420 CACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGG 361

Qy    709 ATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGG 768
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    360 ATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGG 301

Qy    769 GAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACAT 828
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    300 GAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACAT 241

Qy    829 CCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCT 888
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    240 CCCAAGGCTAACCTAACGTACTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCT 181

Qy    889 TCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGA 948
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 TCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGA 121

Qy    949 GAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCC 1008
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    120 GAATCCTTAAAAAGACAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCC 61

Qy   1009 ATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATA 1054
          |||||||||||||||||||||||||||||||||||||||||||||
Db     60 ATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATA 15


RESULT 22
BF997044
LOCUS       BF997044                591 bp    mRNA    linear   EST 23-JAN-2001
DEFINITION  QV3-GN0201-081100-437-b09 GN0201 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF997044
VERSION     BF997044.1  GI:12403367
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV3&t2=QV3-GN0201-
```

```
           081100-437-b09&t3=2000-11-08&t4=1)
           Seq primer: puc 18 forward
           High quality sequence stop: 543.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="GN0201"
                     /dev_stage="Adult"
                     /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                     ; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
BASE COUNT      151 a    116 c    146 g    178 t
ORIGIN


  Query Match             51.3%;  Score 565.6;  DB 10;  Length 591;
  Best Local Similarity   97.6%;  Pred. No. 1.5e-100;
  Matches  574;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy    425 AAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTT 484
          | || ||||        ||||||||||||||||||||||||||||||||||||||||||||
Db      3 AGGTTCTCGACTCGTTACAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTT 62

Qy    485 CCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATC 544
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63 CCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATC 122

Qy    545 TTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATG 604
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    123 TTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATG 182

Qy    605 TAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCC 664
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 TAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCC 242

Qy    665 CTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGC 724
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243 CTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGC 302

Qy    725 TTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAA 784
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 TTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAA 362

Qy    785 ATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAA 844
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363 ATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAA 422

Qy    845 CGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCC 904
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    423 CGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCC 482

Qy    905 CTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAA 964
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    483 CTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAA 542

Qy    965 AAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTA 1012
          |||||||||||||| |||||||||||| | |||||| ||||||||
Db    543 AAATTGACTTGCTTGGCTAAAACTACAGTGGCGCCTTAGGCTCCATTA 590


RESULT  23
BF852437
LOCUS       BF852437                 638 bp    mRNA    linear   EST 16-JAN-2001
DEFINITION  MR3-EN0087-121200-005-a02 EN0087 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF852437
VERSION     BF852437.1  GI:12239808
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 638)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR3&t2=MR3-EN0087-
            121200-005-a02&t3=2000-12-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 16
            High quality sequence stop: 452.
FEATURES             Location/Qualifiers
     source          1. .638
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0087"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      155 a    135 c    161 g    186 t      1 others
ORIGIN


  Query Match             51.2%;  Score 564.6;  DB 10;  Length 638;
  Best Local Similarity   96.6%;  Pred. No. 2.3e-100;
  Matches  576;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy    451 TGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATT 510
          ||  |||  ||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     43 TGGCGGGGTACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACGGGACGTTTATATATT 102

Qy    511 ATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGAC 570
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    103 ATAGGCAGGGCTGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGAC 162

Qy    571 TGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTA 630
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    163 TGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTA 222

Qy    631 TTTATTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAAT 690
          |||||||||||||| |||||||||||||||||||||||||||||||||||| ||||||
Db    223 TTTATTTCGTTAGATGGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCCAGAAAT 282

Qy    691 AGGAAAGGCTAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGA 750
          ||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||
Db    283 AGGAAAGGCTAGGATGGGATAATGCTGCAAATGCTTTTTTGGTGTGAGAAACTGGAGAGA 342

Qy    751 TGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATA 810
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    343 TGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATA 402
```

```
Qy     811 GTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATG 870
           ||| ||||||||||||||||||||||||||||||||||||||||||| ||||||| |||||
Db     403 GTTGGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAACGGTAGATGGAATG 462

Qy     871 AGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTC 930
           |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db     463 CGACATGTTTTCTGTGCTTCTAAGGGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTC 522

Qy     931 CATTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTAC 990
           |||||||||||||||||||||||||||||| |||||||||||| |||||||||||| |||
Db     523 CATTATGGCAGGTTTAGAGAATCCTTAAAAGGAAAAATTGACTGGCTTGCCTAAAAGTAC 582

Qy     991 AGTGCCCCCTTAGCCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAA 1046
           |||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||
Db     583 AGTGCCCCCTTAGCCTCCATTACTTAGTATCTCTTACAGTTTGGTCTGGNTCTCAA 638
```

RESULT 24
BF852557/c

LOCUS	BF852557	634 bp	mRNA	linear	EST 16-JAN-2001

DEFINITION MR3-EN0087-131200-010-e01 EN0087 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF852557
VERSION BF852557.1 GI:12240146
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 634)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR3&t2=MR3-EN0087-131200-010-e01&t3=2000-12-13&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 48
High quality sequence stop: 583.

FEATURES Location/Qualifiers
source 1. .634
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="EN0087"
/dev_stage="Adult"
/note="Organ: lung_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 188 a 151 c 139 g 156 t
ORIGIN

Query Match 50.9%; Score 561.6; DB 10; Length 634;
Best Local Similarity 95.9%; Pred. No. 8.8e-100;
Matches 587; Conservative 0; Mismatches 24; Indels 1; Gaps 1;

```
Qy     442 CAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGT 501
           |||||||| ||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     634 CAGCAAGCCTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCGACCACTGGACGT 575

Qy     502 TTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCC 561
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     574 TTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTCGCTAGCC 515

Qy     562 AGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCT 621
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     514 AGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCT 455

Qy     622 CTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGG 681
           |||||||||||||||||| |||| ||||||||||||||||||||||||||||||||||||
Db     454 CTAAGTTTATTTATTTCGCTAGAGTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGG 395

Qy     682 CCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAA 741
           || |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     394 CCCAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCACATGCTTTTTTTGTGTGAGAAA 335

Qy     742 CTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCC 801
           |||||||||||||||||||||||||||||||||||||||||||| |||||||||| |||||||
Db     334 CTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCACAATTGATTAGCCTGAGCC 275

Qy     802 TTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTA 861
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     274 TTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGCTA 215

Qy     862 GATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCT 921
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     214 GATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCT 155

Qy     922 TCATGTTTCCATTATGGCAGGTTTAGAGAATCC-TTAAAAAGAAAAATTGACTTGCTTGC 980
           |||||||||||||||||| |||||||||||||| ||| ||||||||||||||||||||||
Db     154 TCATGTTTCCATTATGGCCGGTTTAGAGAATCCTTTACAAAGAAAAATTGACTTGCTTGC 95

Qy     981 CTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGC 1040
           |||||||||||||||||||||||||||||| |||||||||||||||||||||      ||
Db      94 CTAAAACTACAGTGCCCCCTTAGCCTCCAGTACTTAGTATCTCTTACAGTTTCGCTCTGC 35

Qy    1041 TCTCAAATAATA 1052
           | | | |  |||
Db      34 TGTAATAATATA 23
```

RESULT 25
BF768839/c

LOCUS	BF768839	588 bp	mRNA	linear	EST 12-JAN-2001

DEFINITION PM0-IT0018-091100-001-h11 IT0018 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF768839
VERSION BF768839.1 GI:12116843
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 588)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research

```
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
          Brazil
          Tel: +55-11-2704922
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM0&t2=PM0-IT0018-
          091100-001-h11&t3=2000-11-09&t4=1)
          Seq primer: puc 18 forward
          High quality sequence start: 18
          High quality sequence stop: 588.
FEATURES             Location/Qualifiers
     source          1. .588
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="IT0018"
                     /dev_stage="Adult"
                     /note="Organ: epid_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      181 a    139 c    111 g    157 t
ORIGIN


  Query Match             50.8%;  Score 560.2;  DB 10;  Length 588;
  Best Local Similarity   99.5%;  Pred. No. 1.7e-99;
  Matches  562;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    485 CCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATC 544
          ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db    588 CCTCAACCACTGGACGTTTATATATCATAGGCAGGGATGAAATGGGCACCTAGTCAGATC 529

Qy    545 TTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATG 604
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    528 TTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATG 469

Qy    605 TAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCC 664
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    468 TAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCC 409

Qy    665 CTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGC 724
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    408 CTGTAGTCCAGGTAGGGCCCAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGC 349

Qy    725 TTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAA 784
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    348 TTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAA 289

Qy    785 ATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAA 844
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    288 ATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTTACCTAA 229

Qy    845 CGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCC 904
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    228 CGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCC 169

Qy    905 CTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAA 964
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    168 CTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAA 109

Qy    965 AAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTATCTCT 1024
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    108 AAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTATCTCT 49

Qy   1025 TACAGTTTGCTCTGGCTCTCAAATA 1049
          |||||||||||||||||||||||||
Db     48 TACAGTTTGCTCTGGCTCTCAAATA 24
```

```
RESULT  26
AK011334
LOCUS       AK011334               1070 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:2610007F17:endosulfine alpha, full insert
            sequence.
ACCESSION   AK011334
VERSION     AK011334.1  GI:12847390
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 10 days embryo cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2610007F17.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1070)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
```

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 7.5 and subtraction to Rot = 37.5. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

```
FEATURES             Location/Qualifiers
     source          1. .1070
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGD:MGI:1894066"
                     /db_xref="taxon:10090"
                     /clone="2610007F17"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days embryo"
     gene            142. .495
                     /gene="Ensa"
     CDS             142. .495
                     /gene="Ensa"
                     /note="data source:MGD, source key:MGI:1891189,
                     evidence:ISS
                     endosulfine alpha
                     putative"
                     /codon_start=1
                     /protein_id="BAB27552.1"
                     /db_xref="GI:12847391"
                     /translation="MSQKQEEENPAEETGEEKQDTQEEEGILPEKAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGADKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAG"
BASE COUNT      273 a    274 c    291 g    232 t
ORIGIN
```

Query Match 50.6%; Score 557.6; DB 11; Length 1070;
Best Local Similarity 78.0%; Pred. No. 4.8e-99;
Matches 839; Conservative 0; Mismatches 189; Indels 48; Gaps 12;

```
Qy       1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
Db      38 GTGACAGGAGCCGGAGCAGCAGCTCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTTTC 97

Qy      61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
Db      98 CGGCTGACTTCCCGGGCCC--TGCATTACACAGTCCCGGTTCTGCCATGTCCCAGAAACA 155

Qy     121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
Db     156 AGAAGAAGAAAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGATACACAGGAGGAAGA 215

Qy     181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
Db     216 AGGGATTCTCCCTGAGAAAGCTGAGGAGGCAAAGCTAAAGGCCAAATATCCAAGCCTAGG 275

Qy     241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
Db     276 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 335

Qy     301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
Db     336 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAACAAGCAGCTGCCAAGTGC 395

Qy     361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
Db     396 AGGAGCAGACAAGAACCTGGTGACCGGTGACCACATCCCCACCCCACAGGACCTGCCCCA 455

Qy     421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCC 480
Db     456 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGGG--CCCCTCTCTTCC 513

Qy     481 CCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCA 540
Db     514 CCTTCCTCACCCACTGGACTGTTATATATTGTAGGCAGGG----AATGGGCACCTAGTCA 569

Qy     541 GATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGT 600
Db     570 GAGCTCA-------GCTTGCCATCCAGGGGAGACATTCTGCTGGGGGCT---------- 611

Qy     601 AATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACA 660
Db     612 ----TAGGTTGAAGATGGGCTCTGAGTTCATTTCTTTGGGTAAATTGAAACTTGTACACC 667

Qy     661 CTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAA 720
Db     668 CTCATGATAGCCCAGTAGAGACCCAAGAGTAGGAAAGACTAGGAGTGCAGAGTGCTGCA- 726

Qy     721 ATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCC 780
Db     727 --GCTTTTCCTGTGGTGAGAATGTGGAAGATGCAAGTTCTCCTGTTGGAGGAGCAAGTCC 784

Qy     781 CAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAAC 840
Db     785 TTAGACGGGCTATCGT--------GGAGCAGTGTGACAGGTTCAGTGTCCCAAGGCTGAA 836

Qy     841 CTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCT 900
Db     837 TTGACAT-GCTGAGAAAGGTAGAGTGAAGGAG--ATGCTCTCTGTGCTTCTAAGTGTGCT 893

Qy     901 GTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAA 960
Db     894 GTCCCTTAGGC--CTCTGTGCTCTGTGTTCCCATCGTGGGTGTTTAGGGGACCTTCAAAA 951

Qy     961 AGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTAT 1020
Db     952 GAGAGGAGATTTCTGCTTGCTTAAAATTACAGGGCCCACCTAACC-CCTCTGCTGTGTTC 1010

Qy    1021 CTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACATTATTCAC 1076
Db    1011 CTCCAACAG-CTGCCCTGCTTCTCAAGGAGTGTAAAGTTTAACAGACGTTAGCTAC 1065
```

```
RESULT 27
AK019170
LOCUS       AK019170                1067 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:2610104C01:endosulfine alpha, full insert
            sequence.
ACCESSION   AK019170
VERSION     AK019170.1  GI:12859238
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 10 days embryo cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2610104C01.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
```

```
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1067)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .1067
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGD:MGI:1899583"
                     /db_xref="taxon:10090"
                     /clone="2610104C01"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days embryo"
     gene            142. .495
                     /gene="Ensa"
     CDS             142. .495
                     /gene="Ensa"
                     /note="data source:MGD, source key:MGI:1891189,
                     evidence:ISS
                     endosulfine alpha
                     putative"
                     /codon_start=1
                     /protein_id="BAB31584.1"
                     /db_xref="GI:12859239"
                     /translation="MSQKQEEENPAEETGEEKQDTQEEEGILPEKAEEAKLKAKYPSL
                     GQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGADKNLVTGDHIPTPQD
                     LPQRKSSLVTSKLAG"
BASE COUNT      271 a    271 c    290 g    235 t
ORIGIN


  Query Match           50.4%;  Score 556;  DB 11;  Length 1067;
  Best Local Similarity 77.9%;  Pred. No. 9.8e-99;

  Matches  838;  Conservative    0;  Mismatches  190;  Indels   48;  Gaps   12;

Qy      1 GTGACAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTC 60
          ||||||||||||| |||||||| ||||||||||||||||||||||||||||||||||| ||
Db     38 GTGACAGGAGCCGGAGCAGCAGCTCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTTTC 97

Qy     61 CGGTTGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACA 120
          ||| || |||||||||||||  | || | |||||||||||| | ||||||||||||||||
Db     98 CGGCTGACTTCCCGGGCCC--TGCATTACACAGTCCCGGTTCTGCCATGTCCCAGAAACA 155

Qy    121 AGAAGAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGA 180
          |||||||| |||||||||||||||||||||||||||||||||| || |||||| ||||
Db    156 AGAAGAAGAAAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGATACACAGGAGGAAGA 215

Qy    181 AGGTATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGG 240
          ||| ||||| |||||| |||||| ||||||||||||||||||||| ||||||||||
Db    216 AGGGATTCTCCCTGAGAAAGCTGAGGAGGCAAAGCTAAAGGCCAAATATCCAAGCCTAGG 275

Qy    241 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    276 ACAAAAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTT 335

Qy    301 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGC 360
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    336 TGACTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAACAAGCAGCTGCCAAGTGC 395

Qy    361 AGGACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCA 420
          |||| |||||||||||||||||| ||||| |||||||||||||||||| ||||||||
Db    396 AGGAGCAGACAAGAACCTGGTGACCGGTGACCACATCCCCACCCCACAGGACCTGCCCCA 455

Qy    421 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCC 480
          |||||||||||||||||||||||||||||||||||||||||||| |  |||||||| |||
Db    456 GAGAAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGGG--CCCCTCTCTTCC 513

Qy    481 CCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCA 540
          |||||||| |||||||| |||||||| |||||||| |||||||||||||||||||||||
Db    514 CCTTCCTCACCCACTGGACTGTTATATATTGTAGGCAGGG----AATGGGCACCTAGTCA 569

Qy    541 GATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGT 600
          || ||         ||| ||||   | |   |  ||| ||||||||||
Db    570 GAGCTCA-------GCTTGCCATCCAGGGGAGACATTTTGCTGGGGGCT---------- 611

Qy    601 AATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACA 660
            |||||||| | ||||||| |||| |||| ||| | || ||||| ||||  ||||
Db    612 ----TAGGTTGAAGATGGGCTCTGAGTTCATTTCTTTGGGTAAATTGAAACTTGTACACC 667

Qy    661 CTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAA 720
          |||    ||| ||||    | || |  | |||||||| |||||| || | | |||||||
Db    668 CTCATGATAGCCCAGTAGAGACCCAAGAGTAGGAAAGACTAGGAGTGCAGAGTGCTGCA- 726

Qy    721 ATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCC 780
            ||||||  ||||      ||   |  |||||  | |||||||  |||| |||  | ||||
Db    727 --GCTTTTCCTGTGGTGAGAATGTGGAAGATGCAAGTTCTCCTGTTGGAGGAGCAAGTCC 784

Qy    781 CAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAAC 840
            | |  |  ||   |         |||  ||| || ||||||  |   ||||||||| |
Db    785 TTAGACGGGCTATCGT--------GGAGCAGTGTGACAGGTTCAGTGTCCCAAGGCTGAA 836

Qy    841 CTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCT 900
            | || | | || |||||||||| |||| |||  ||| | |||||||| ||||||| ||
Db    837 TTGACAT-GCTGAGAAAGGTAGAGTGAAGGAG--ATGCTCTCTGTGCTTTTAAGTGTGCT 893

Qy    901 GTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAA 960
          ||||||||||||  || |    ||    ||||  |   |   |||||| | | | | ||||
Db    894 GTCCCTTAGGCCTCTGTGC--TCTGTGTTCCCATCGTGGGTGTTTAGGGGACCTTCAAAA 951

Qy    961 AGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTAT 1020
            | |      ||||||| ||||| ||||| |||||| || || || || ||  ||
Db    952 GAGAGGAGATTTCTGCTTGCTTAAAATTACAGGGCCCCCCTAACC-CCTTTGCTGTGTTC 1010

Qy   1021 CTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACATTATTCAC 1076
          |||  ||||  ||| |||  | ||||  | | ||||| || |   || |||   ||
Db   1011 CTCCAACAG-CTGCCCTGCTTTTCAAGGAGTGTAAAGTTTAACAGACGTTAGCTAC 1065
```

```
RESULT  28
BF852494
LOCUS       BF852494                612 bp    mRNA    linear   EST 16-JAN-2001
DEFINITION  MR3-EN0087-121200-006-g12 EN0087 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF852494
VERSION     BF852494.1  GI:12239869
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 612)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR3&t2=MR3-EN0087-
            121200-006-g12&t3=2000-12-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 8
            High quality sequence stop: 529.
FEATURES             Location/Qualifiers
     source          1. .612
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0087"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      158 a    124 c    157 g    173 t
ORIGIN

  Query Match           50.3%;  Score 555;  DB 10;  Length 612;
  Best Local Similarity   96.5%;  Pred. No. 1.7e-98;
  Matches  578;  Conservative    0;  Mismatches   20;  Indels    1;  Gaps    1;

Qy     455 GGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAG 514
           |||  |||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db      15 GGGGTACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACGGGACGTTTATATATTATAG 74

Qy     515 GCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTG 574
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75 GCA-GGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTG 133

Qy     575 ATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTA 634
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     134 ATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTA 193

Qy     635 TTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGA 694
           |||||||||| |||||||||||||||||||| |||||||||||||||| |||||||||
Db     194 TTTCGTTAGATGGACACTTCCACACACTCCCGGTAGTCCAGGTAGGGCCCAGAAATAGGA 253

Qy     695 AAGGCTAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTG 754
           |||||||||| ||||||||||||||||| ||||||| |||||||||||||||||||||||
Db     254 AAGGCTAGGATGGGATAATGCTGCAAATACTTTTTTGGTGTGAGAAACTGGAGAGATGTG 313

Qy     755 ATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTT 814
           |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db     314 ATTTCTCCTTTTGGGAGAGAATGTCCCAAAATGGATTAGGCTGAGCCTTGGGAATAGTTT 373

Qy     815 GGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGAC 874
           |||||||||||||||||||||||||||||||||| |||| |||||||||||| |||||||||
Db     374 GGCAGGTTTAACATCCCAAGGCTAACCTAAGGTAGTGGGGAAAGGTAGATGGAATGAGAC 433

Qy     875 ATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATT 934
           |||||||||||||||||| |||||||||||||||||||| |||||||||||||||||||
Db     434 ATGTTTTCTGTGCTTCTAAGGGTTCTGTCCCTTAGGCTGCTATGGCTTCATGTTTCCATT 493

Qy     935 ATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTG 994
           |||||||||||||||||||||||||||||||||| |||| ||| ||||||||||||||||
Db     494 ATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATGGACTGGCTGGCCTAAAACTACAGTG 553

Qy     995 CCCCCTTAGCCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATAT 1053
           |||||||||||||||||||||||||||||||||||||||||||| || ||||||||||||
Db     554 CCCCCTTAGCCTCCATTACTTAGTATCTCTTACAGTTTGCTCGGGGTCTCAAATAATAT 612
```

```
RESULT  29
BM007320
LOCUS       BM007320                689 bp    mRNA    linear   EST 30-OCT-2001
DEFINITION  603616529F1 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:5440590 5',
            mRNA sequence.
ACCESSION   BM007320
VERSION     BM007320.1  GI:16521674
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Mark Watson
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1913  row: i  column: 07
            High quality sequence stop: 636.
FEATURES             Location/Qualifiers
     source          1. .689
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5440590"
                     /clone_lib="NIH_MGC_113"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT      191 a    194 c    175 g    129 t
ORIGIN
```

```
  Query Match             50.0%;  Score 551.6;  DB 10;  Length 689;
  Best Local Similarity   92.2%;  Pred. No. 7.7e-98;
  Matches  618;  Conservative    0;  Mismatches   39;  Indels   13;  Gaps    3;

Qy       5 CAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGT 64
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 CAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGT 61

Qy      65 TGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAA 124
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 TGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAA 121

Qy     125 GAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGT 184
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 GAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGT 181

Qy     185 ATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAA 244
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 ATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAA 241

Qy     245 AAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGAC 304
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 AAGCCTGGAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGAC 301

Qy     305 TCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGA 364
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 TCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGA 361

Qy     365 CCAGACAAGAACCTGGTGACTGGTGATCACATCCCCA-CCCCACAGGATCTGCCCCAGAG 423
           |||||||||||||||||||||||||||||| |||| ||||||||||     || ||
Db     362 CCAGACAAGAACCTGGTGACTGGTGATCACATTCCCAGCCCCACAGGATTCTGGCCCCAG 421

Qy     424 AAAG----------TCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCC 472
           |||          || |   |     |    ||||||||||||||||||||||||||
Db     422 GAAGAAAAGTTCCCTTCGTTCGTTCAACCAGCAAAGCTTGCGGGGTAACCTGAGCCCCCC 481

Qy     473 TCTCCTCCC-CTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGC 531
           ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 TCTCCTCCCGCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGC 541

Qy     532 ACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGC 591
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 ACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGC 601

Qy     592 TGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACAC 651
           |||||||||||||||| |||||||||||||||||||| ||||  |   || |    |
Db     602 TGAGAAGGTAATGTAGGTGGAAAAGGGGCTCTAAGTTTATATATTCAGAAGGACTCGACC 661

Qy     652 TTCCACACAC 661
           |||||
Db     662 ACTTCCACAC 671


RESULT 30
BC020533
LOCUS       BC020533               2256 bp    mRNA    linear   HTC 22-JAN-2002
DEFINITION  Mus musculus, Similar to endosulfine alpha, clone IMAGE:4221098,
            mRNA.
ACCESSION   BC020533
VERSION     BC020533.1  GI:18088212
KEYWORDS    HTC.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2256)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 25 Row: f Column: 3
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, GenomeScan gene prediction, Similarity but not identity
            to protein
            This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1. .2256
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4221098"
                     /tissue_type="Kidney, normal. 5 month old male mouse."
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
BASE COUNT      587 a    518 c    589 g    562 t
ORIGIN


  Query Match             49.8%;  Score 549.6;  DB 11;  Length 2256;
  Best Local Similarity   77.7%;  Pred. No. 1.5e-97;
  Matches  845;  Conservative    0;  Mismatches  194;  Indels   49;  Gaps   13;

Qy       5 CAGGAGCCGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGT 64
           | ||||||| |||||||| ||||||||||||||||||||||||||||||||| |||||
Db  .   11 CGGGAGCCGGAGCAGCAGCTCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTTTCCGGC 70

Qy      65 TGCCTTCCCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAA 124
           || ||||||||||  | || | |||||||||||| | ||||||||||||||||||||||
Db      71 TGACTTCCCGGGCCC--TGCATTACACAGTCCCGGTTCTGCCATGTCCCAGAAACAAGAA 128

Qy     125 GAAGAGAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGT 184
           |||||  |||||||||||||||||||||||||||||||||| || ||||||||||||||
Db     129 GAAGAAAACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGATACACAGGAGAAAGAAGGG 188

Qy     185 ATTCTGCCTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAA 244
           |||||  |||||||  ||||||  |||||||||||||||||||||  ||||||||||||||
Db     189 ATTCTCCCTGAGAAAGCTGAGGAGGCAAAGCTAAAGGCCAAATATCCAAGCCTAGGACAA 248

Qy     245 AAGCCTGGAG-GCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGA 303
           ||||||||||  |||||||||||||||||||||||||||||||||||||||||||||||
Db     249 AAGCCTGGAGAGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGA 308

Qy     304 CTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGG 363
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     309 CTCAGGAGACTACAACATGGCCAAAGCCAAGATGAAGAACAAGCAGCTGCCAAGTGCAGG 368

Qy     364 ACCAGACAAGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAG 423
           | |||||||||||||||||  ||||| ||||||||||||||||||| ||||||||||
Db     369 AGCAGACAAGAACCTGGTGACCGGTGACCACATCCCCACCCCACAGGACCTGCCCCAGAG 428

Qy     424 AAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCT 483
           |||||||||||||||||||||||||||||||||||||||||||||| |  |||||||| ||||||
Db     429 AAAGTCCTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGGG--CCCCTCTCTTCCCCT 486

Qy     484 TCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGAT 543
           ||||||  |||||||||  |||||||||  |||||||||     ||||||||||||||||||
Db     487 TCCTCACCCACTGGACTGTTATATATTGTAGGCAGGG----AATGGGCACCTAGTCAGAG 542
```

```
Qy    544 CTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAAT 603
          ||         ||| ||||   | |   | ||||||||||||||||
Db    543 CT-------CAGCTTGCCATCCAGGGGAGACATTCTGCTGGGGGCT-------------- 581

Qy    604 GTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTC 663
          |||||||||| | |||||||| |||| |||| ||| | || ||||| |||| |||| |||
Db    582 -TAGGTTGAAGATGGGCTCTGAGTTCATTTCTTTGGGTAAATTGAAACTTGTACACCCTC 640

Qy    664 CCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATG 723
            ||| ||||    | || |  | |||||||| |||||| || | | |||||||   |
Db    641 ATGATAGCCCAGTAGAGACCCAAGAGTAGGAAAGACTAGGAGTGCAGAGTGCTGCA---G 697

Qy    724 CTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAA 783
          |||||  ||||     ||   |  |||||  | ||||||| ||||  ||| | ||||  |
Db    698 CTTTTCCTGTGGTGAGAATGTGGAAGATGCAAGTTCTCCTGTTGGAGGAGCAAGTCCTTA 757

Qy    784 AATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTA 843
            |  ||   |          ||| ||| || |||||| |   |||||||||| |  |
Db    758 GACGGGCTATCGT--------GGAGCAGTGTGACAGGTTCAGTGTCCCAAGGCTGAATTG 809

Qy    844 ACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTC 903
          || | | || |||||||||| |||| |||  ||| | ||||||||||||||||| |||||
Db    810 ACAT-GCTGAGAAAGGTAGAGTGAAGGAG--ATGCTCTCTGTGCTTCTAAGTGTGCTGTC 866

Qy    904 CCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGA 963
          ||||||||  || |    ||    |||| |   |   |||||| | | | | ||||
Db    867 CCTTAGGCCTCTGTGC--TCTGTGTTCCCATCGTGGGTGTTTAGGGGACCTTCAAAAGAG 924

Qy    964 AAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTATCTC 1023
          |  |      ||||||| ||||| ||||| |||| | || || ||  | ||  ||  |||
Db    925 AGGAGATTTCTGCTTGCTTAAAATTACAGGGCCCACCTAACC-CCTCTGCTGTGTTCCTC 983

Qy   1024 TTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAACATTATTCACAAAAAAA 1083
            ||||  ||| |||  ||||||  | | ||||| || |   || |||   ||  |  ||
Db    984 CAACAG-CTGCCCTGCTTCTCAAGGAGTGTAAAGTTTAACAGACGTTAGCTACTTATGAA 1042

Qy   1084 AAAAAAAA 1091
          ||| ||||
Db   1043 AAAGAAAA 1050
```

```
RESULT  31
BF853400/c
LOCUS       BF853400                565 bp    mRNA    linear   EST 16-JAN-2001
DEFINITION  MR3-EN0087-191200-015-e12 EN0087 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF853400
VERSION     BF853400.1  GI:12241144
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 565)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR3&t2=MR3-EN0087-
            191200-015-e12&t3=2000-12-19&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 561.
FEATURES             Location/Qualifiers
     source          1. .565
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0087"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      175 a    132 c    118 g    140 t
ORIGIN
```

```
  Query Match             49.3%;  Score 544.2;  DB 10;  Length 565;
  Best Local Similarity   98.6%;  Pred. No. 2.2e-96;
  Matches  549;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy    469 CCCCTCTCCTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATG 528
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    565 CCCCTCTCCTCCCCTTCCTCCACCACTGGACGTTTATATATTATAGGCAGGGATGAAATG 506

Qy    529 GGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGC 588
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    505 GGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGC 446

Qy    589 TGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGA 648
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    445 TGCTGAGAAGGTGATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGA 386

Qy    649 CACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGG 708
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    385 CACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCCAGAAATAGGAAAGGCTAGGATTGG 326

Qy    709 ATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGG 768
          || ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    325 ATCATGCTGCCAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGG 266

Qy    769 GAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACAT 828
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    265 GAGAGAATGTCCCCAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACAT 206

Qy    829 CCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCT 888
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    205 CCCCAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCT 146

Qy    889 TCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGA 948
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145 TCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGA 86

Qy    949 GAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCC 1008
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 GAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCC 26

Qy   1009 ATTACTTAGTATCTCTT 1025
          |||||||||||||||| |
Db     25 ATTACTTAGTATCTCCT 9
```

```
RESULT  32
BM128047
LOCUS       BM128047                568 bp    mRNA    linear   EST 27-NOV-2001
DEFINITION  if09a04.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA 5', mRNA sequence.
```

```
ACCESSION   BM128047
VERSION     BM128047.1  GI:17122599
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 568)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: if09a04.x1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            High quality sequence stop: 429.
FEATURES             Location/Qualifiers
     source          1. .568
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                     /sex="Both"
                     /tissue_type="Islets of Langerhans"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                     Site_2: Sal 1; Starting library constructed using
                     SuperScript Plasmid Library kit  (Life Technologies). cDNA
                     made by oligo-dT priming. Size-selected by  column
                     fractionation; average insert size 1.08 kb. Library was
                     amplified  once on solid support and plasmid DNA from
                     library was prepared. The  library DNA was normalized by
                     method #4 from Bonaldo, Lennon, and Soares  1996 Genome
                     Research 6:791-806; 0.5 microgram single-stranded library
                     plasmid DNA was mixed with 5 micrograms PCR product
                     representing library  inserts and hybridized to an Ecot of
                     20. Single-stranded (unhybridized)  plasmids were isolated
                     by hydroxyapatite chromatography and used to make  this
                     library."
BASE COUNT      153 a    105 c    132 g    178 t
ORIGIN
```

```
  Query Match             49.0%;  Score 540.8;  DB 10;  Length 568;
  Best Local Similarity   99.5%;  Pred. No. 1e-95;
  Matches  553;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy     528 GGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGG 587
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGG 60

Qy     588 CTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTG 647
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTG 120

Qy     648 ACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTG 707
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     121 ACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCCAGAAATAGGAAAGGCTAGGATTG 180

Qy     708 GATAATGCTGCAAATGC-TTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTT 766
           |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db     181 GATAATGCTGCAAATGCTTTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTT 240

Qy     767 GGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAAC 826
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAAC 300

Qy     827 ATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTG 886
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTG 360

Qy     887 CTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTA 946
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTA 420

Qy     947 GAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCT 1006
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCT 480

Qy    1007 CCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAA 1066
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGATGAA 540

Qy    1067 CATTATTCACAAAAAA 1082
           |||||||||||| ||||
Db     541 CATTATTCACAGAAAA 556
```

```
RESULT  33
BF768834
LOCUS       BF768834                563 bp    mRNA    linear   EST 12-JAN-2001
DEFINITION  PM0-IT0018-091100-001-f12 IT0018 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF768834
VERSION     BF768834.1  GI:12116838
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 563)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM0&t2=PM0-IT0018-
            091100-001-f12&t3=2000-11-09&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 562.
FEATURES             Location/Qualifiers
     source          1. .563
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="IT0018"
                     /dev_stage="Adult"
                     /note="Organ: epid_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
```

derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 144 a 115 c 138 g 166 t
ORIGIN

Query Match 48.9%; Score 539.4; DB 10; Length 563;
Best Local Similarity 99.5%; Pred. No. 1.9e-95;
Matches 562; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

```
Qy    441 CCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACG 500
          ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      1 CCAGCAAGCTTGCGGGGT-ACCTGAGCCCCCCTCTCCTCCCCTTCCTCAACCACTGGACG 59

Qy    501 TTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGC 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60 TTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGC 119

Qy    561 CAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGC 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    120 CAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGC 179

Qy    621 TCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGG 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 TCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGG 239

Qy    681 GCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAA 740
          ||| |||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    240 GCCCAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGC-TTTTTTGTGTGAGAA 298

Qy    741 ACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGC 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299 ACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGC 358

Qy    801 CTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGT 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    359 CTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGT 418

Qy    861 AGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGC 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    419 AGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGC 478

Qy    921 TTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGC 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    479 TTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGC 538

Qy    981 CTAAAACTACAGTGCCCCCTTAGCC 1005
          |||||||||||||||||||||||||
Db    539 CTAAAACTACAGTGCCCCCTTAGCC 563
```

RESULT 34
BF852560/c
LOCUS BF852560 601 bp mRNA linear EST 16-JAN-2001
DEFINITION MR3-EN0087-131200-010-a12 EN0087 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF852560
VERSION BF852560.1 GI:12240154
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 601)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR3&t2=MR3-EN0087-131200-010-a12&t3=2000-12-13&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 19
High quality sequence stop: 599.
FEATURES Location/Qualifiers
source 1. .601
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="EN0087"
/dev_stage="Adult"
/note="Organ: lung_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 182 a 142 c 120 g 157 t
ORIGIN

Query Match 48.8%; Score 538.4; DB 10; Length 601;
Best Local Similarity 97.9%; Pred. No. 3e-95;
Matches 556; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

```
Qy    488 CAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTC 547
          |||||||||||||||||||||||||| |||||||||||||  ||||||||||||||||||
Db    600 CAACCACTGGACGTTTATATATTATAGTCAGGGATGAAATGGCAACCTAGTCAGATCTTC 541

Qy    548 TCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAG 607
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 TCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAG 481

Qy    608 GTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCTG 667
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    480 GTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCTCACACTCCCTG 421

Qy    668 TAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGCTTT 727
          |||||||||||||||| |||||||||||||||||||||||||| ||||||||||||||||
Db    420 TAGTCCAGGTAGGGCCCAGAAATAGGAAAGGCTAGGATTGGATAATCCTGCAAATGCTTT 361

Qy    728 TTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGA-GAGAATGTCCCAAAAT 786
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||  |||
Db    360 TTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGACGAGAATGTCCCCCAAT 301

Qy    787 TGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACG 846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    300 TGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACG 241

Qy    847 TAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCT 906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    240 TAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCT 181

Qy    907 TAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAA 966
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 TAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAA 121

Qy    967 ATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGTATCTCTTA 1026
          |||||||||||||||| |||||||||||||||||||||||| ||||||||||||||||||
```

```
Db    120 ATTGACTTGCTTGCCTACAACTACAGTGCCCCCTTAGCCTCCCTTACTTAGTATCTCTTA 61

Qy   1027 CAGTTTGCTCTGGCTCTCAAATAATATA 1054
          |||||||||||||||||||||||||| |
Db     60 CAGTTTGCTCTGGCTCTCAAATAATAAA 33


RESULT  35
BF773208/c
LOCUS       BF773208                616 bp    mRNA    linear   EST 12-JAN-2001
DEFINITION  PM0-IT0018-151200-003-f11 IT0018 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF773208
VERSION     BF773208.1  GI:12121108
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 616)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM0&t2=PM0-IT0018-
            151200-003-f11&t3=2000-12-15&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 7
            High quality sequence stop: 145.
FEATURES             Location/Qualifiers
     source          1. .616
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="IT0018"
                     /dev_stage="Adult"
                     /note="Organ: epid_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      192 a    141 c    124 g    159 t
ORIGIN


  Query Match             48.5%;  Score 535.2;  DB 10;  Length 616;
  Best Local Similarity   98.1%;  Pred. No. 1.2e-94;
  Matches  573;  Conservative    0;  Mismatches    8;  Indels    3;  Gaps    3;

Qy    480 CCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTC 539
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    607 CCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTC 548

Qy    540 AGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGG 599
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    547 AGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGG 488

Qy    600 TAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACAC 659
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    487 TAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACAC 428

Qy    660 ACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCA 719
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    427 ACTCCCTGTAGTCCAGGTAGGGCCCAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCG 368

Qy    720 AATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTC 779
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    367 AATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTC 308

Qy    780 CCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAA 839
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    307 CCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAA 248

Qy    840 CCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTC 899
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    247 CCTAACGTAGTTGGGAATGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTC 188

Qy    900 TGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGG-TTTAGAGAATCCTTAA 958
          |||||||||||||||||||||||||||||||||| ||||| ||| ||||||||||||
Db    187 TGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATTGCAGGTTTTGGAGAATCCTTAA 128

Qy    959 AAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGT 1018
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    127 AAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCATTACTTAGT 68

Qy   1019 ATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGATTGA 1062
          ||||| |||||||||||||| |||||||||| |||||| ||| ||
Db     67 ATCTC-TACAGTTTGCTCT-GCTCTCAAATTATATAACGATCGA 26


RESULT  36
BF996987/c
LOCUS       BF996987                549 bp    mRNA    linear   EST 23-JAN-2001
DEFINITION  QV3-GN0201-081100-468-f04 GN0201 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF996987
VERSION     BF996987.1  GI:12403310
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV3&t2=QV3-GN0201-
            081100-468-f04&t3=2000-11-08&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 530.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /clone_lib="GN0201"
                     /dev_stage="Adult"
                     /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                     ; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
BASE COUNT      171 a    129 c    101 g    148 t
ORIGIN


  Query Match             48.5%;  Score 535;  DB 10;  Length 549;
  Best Local Similarity   99.1%;  Pred. No. 1.4e-94;
  Matches  538;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy     520 GATGAAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCT 579
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     549 GATGAAATGGGCACTTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCT 490

Qy     580 GCTGGGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCG 639
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     489 GCTGGGGCCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCG 430

Qy     640 TTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGC 699
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     429 TTAGATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCCAGAAATAGGAAAGGC 370

Qy     700 TAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTC 759
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     369 TAGGATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTC 310

Qy     760 TCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAG 819
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     309 TCCTTTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAG 250

Qy     820 GTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTT 879
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     249 GTTTAACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTT 190

Qy     880 TTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGC 939
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189 TTCTGTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGC 130

Qy     940 AGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCC 999
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     129 AGGTTTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCC 70

Qy    1000 TTAGCCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGAT 1059
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db      69 TTAGCCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAACGAT 10

Qy    1060 TGA 1062
           ||
Db       9 CGA 7


RESULT  37
BF996978/c
LOCUS       BF996978                543 bp    mRNA    linear   EST 23-JAN-2001
DEFINITION  QV3-GN0201-081100-468-c05 GN0201 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF996978
VERSION     BF996978.1  GI:12403301
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV3&t2=QV3-GN0201-
            081100-468-c05&t3=2000-11-08&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 543.
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="GN0201"
                     /dev_stage="Adult"
                     /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                     ; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
BASE COUNT      170 a    126 c    100 g    146 t      1 others
ORIGIN


  Query Match             48.2%;  Score 531.8;  DB 10;  Length 543;
  Best Local Similarity   99.4%;  Pred. No. 5.9e-94;
  Matches  533;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy     524 AAATGGGCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTG 583
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     543 AAATGGCCACCTAGTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTG 484

Qy     584 GGGGCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAG 643
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     483 GGGNCTGCTGAGAAGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAG 424

Qy     644 ATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGG 703
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     423 ATTGACACTTCCACACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGG 364

Qy     704 ATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCT 763
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     363 ATTGGATAATGCTGCAAATGCTTTTTTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCT 304

Qy     764 TTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTT 823
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     303 TTTGGGAGAGAATGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTT 244

Qy     824 AACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCT 883
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     243 AACATCCCAAGGCTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCT 184

Qy     884 GTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGT 943
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     183 GTGCTTCTAAGTGTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGT 124

Qy     944 TTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAG 1003
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123 TTAGAGAATCCTTAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAG 64
```

```
Qy   1004 CCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAAAGAT 1059
          ||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     63 CCTCCATTACTTAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAATATAACGAT 8


RESULT 38
BF997034/c
LOCUS       BF997034                576 bp    mRNA    linear   EST 23-JAN-2001
DEFINITION  QV3-GN0201-081100-437-b11 GN0201 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF997034
VERSION     BF997034.1  GI:12403357
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV3&t2=QV3-GN0201-
            081100-437-b11&t3=2000-11-08&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 576.
FEATURES             Location/Qualifiers
     source          1. .576
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="GN0201"
                     /dev_stage="Adult"
                     /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                     ; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
BASE COUNT      179 a    129 c    122 g    145 t      1 others
ORIGIN


  Query Match             48.2%;  Score 531.8;  DB 10;  Length 576;
  Best Local Similarity   97.4%;  Pred. No. 5.8e-94;
  Matches  561;  Conservative    0;  Mismatches   13;  Indels    2;  Gaps    2;

Qy    477 CTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTA 536
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    576 CTCCCCTTCCTCAACCACTGGACGTTTATATATTATAGNCAGGGATGAAATGGGCACCTA 517

Qy    537 GTCAGATCTTCTCAGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGA 596
          |||||||||||||||||||||||||| |||||||||||||||||||| |||||||||
Db    516 GTCAGATCTTCTCAGCTTGCTAGCCAGCAATGACTGTGATTCTGCTGGGGTCTGCTGAGA 457

Qy    597 AGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCA 656
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    456 AGGTAATGTAGGTTGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCA 397

Qy    657 CACACTCCCTGTAGTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCT 716
          |||||||||||||||||||||||||| ||| ||||||||||||||||||||||||| ||
Db    396 CACACTCCCTGTAGTCCAGGTAGGGCCCAGACATAGGAAAGGCTAGGATTGGATAATCCT 337

Qy    717 GCAAATGCTTTT-TTTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAA 775
          |||||||||||| |||||||||| ||||||||||||||||||||||||||||||||||||
Db    336 GCAAATGCTTTTCTTTGTGTGAGACACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAA 277

Qy    776 TGTCCCAAAATTGATTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGG 835
          |||||||||| ||||||| ||||||||||||||||||||||||||||||||||||||||
Db    276 TGTCCCAAAACTGATTAGCCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGG 217

Qy    836 CTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGT 895
          |||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||
Db    216 CTAACCTAACGTAGTTGGGAAAGGTAGATTGAATGAGACATGTCTTCTCTGCTTCTAAGT 157

Qy    896 GTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCT 955
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    156 GTTCTGTCCCTTAGGCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCT 97

Qy    956 TAAAAAGAAAAATTGACTTGCTTGCCTAAAACTACAGTGCCCCCTTAGCCTCCA-TTACT 1014
          |||||||||||| |||||||||||||||||||| |||||||||||||||| |||||
Db     96 TAAAAAGAAAACTTGACTTGCTTGCCTAAAACTACCGTGCCCCCTTAGCCTCCATTTACT 37

Qy   1015 TAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAA 1050
          |||||||||||||||||||||||||||||||||||
Db     36 TAGTATCTCTTACAGTTTGCTCTGGCTCTCAAATAA 1


RESULT 39
BI258970
LOCUS       BI258970                725 bp    mRNA    linear   EST 17-JUL-2001
DEFINITION  602971620F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5111069 5',
            mRNA sequence.
ACCESSION   BI258970
VERSION     BI258970.1  GI:14815847
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 725)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11269  row: g  column: 06
            High quality sequence stop: 575.
FEATURES             Location/Qualifiers
     source          1. .725
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5111069"
                     /clone_lib="NIH_MGC_12"
                     /tissue_type="cervical carcinoma cell line"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.4 kb. Library prepared by Life
                     Technologies."
BASE COUNT      210 a    206 c    177 g    131 t      1 others
ORIGIN
```

```
  Query Match             48.2%;  Score 531.4;  DB 10;  Length 725;
  Best Local Similarity   99.5%;  Pred. No. 6.7e-94;
  Matches  554;  Conservative    0;  Mismatches    1;  Indels    2;  Gaps    2;

Qy      12 CGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTC 71
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CGAAGCAGCAGCGCAGGTTGTCCCCGTTTCCCCTCCCCCTTCCCTTCTCCGGTTGCCTTC 60

Qy      72 CCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGA 131
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCGGGCCCCTTACACTCCACAGTCCCGGTCCCGCCATGTCCCAGAAACAAGAAGAAGAGA 120

Qy     132 ACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGC 191
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ACCCTGCGGAGGAGACCGGCGAGGAGAAGCAGGACACGCAGGAGAAAGAAGGTATTCTGC 180

Qy     192 CTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTG 251
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTGAGAGAGCTGAAGAGGCAAAGCTAAAGGCCAAATACCCAAGCCTAGGACAAAAGCCTG 240

Qy     252 GAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAG 311
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GAGGCTCCGACTTCCTCATGAAGAGACTCCAGAAAGGGCAAAAGTACTTTGACTCAGGAG 300

Qy     312 ACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACA 371
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ACTACAACATGGCCAAAGCCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGACA 360

Qy     372 AGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCT 431
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AGAACCTGGTGACTGGTGATCACATCCCCACCCCACAGGATCTGCCCCAGAGAAAGTCCT 420

Qy     432 CGCTCGTCACCAGCAAGCTTG-CGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAA 490
           ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     421 CGCTCGTCACCAGCAAGCTTGCCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCAA 480

Qy     491 CCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGC-ACCTAGTCAGATCTTCTC 549
           |||||||||||| |||||||||||||||||||||||||||| |||||||||||||||||||
Db     481 CCACTGGACGTATATATATTATAGGCAGGGATGAAATGGGCAACCTAGTCAGATCTTCTC 540

Qy     550 AGCTTGCTAGCCAGAAA 566
           |||||||||||||||||
Db     541 AGCTTGCTAGCCAGAAA 557


RESULT  40
BF852439
LOCUS       BF852439                 578 bp    mRNA    linear   EST 16-JAN-2001
DEFINITION  MR3-EN0087-121200-005-c01 EN0087 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF852439
VERSION     BF852439.1  GI:12239810
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 578)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR3&t2=MR3-EN0087-
            121200-005-c01&t3=2000-12-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 578.
FEATURES             Location/Qualifiers
     source          1. .578
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0087"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      148 a    112 c    154 g    164 t
ORIGIN


  Query Match             47.5%;  Score 524.4;  DB 10;  Length 578;
  Best Local Similarity   96.0%;  Pred. No. 1.6e-92;
  Matches  549;  Conservative    0;  Mismatches   21;  Indels    2;  Gaps    1;

Qy     430 CTCGCTCGTCACCAGCAAGCTTGCGGGGTAACCTGAGCCCCCCTCTCCTCCCCTTCCTCA 489
           |||| |||| |||||||| | |||||||||| ||||||||||||||||||||||||||||
Db       6 CTCGATCGTTACCAGCAA--TGGCGGGGTAACTTGAGCCCCCCTCTCCTCCCCTTCCTCA 63

Qy     490 ACCACTGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTC 549
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64 ACCACGGGACGTTTATATATTATAGGCAGGGATGAAATGGGCACCTAGTCAGATCTTCTC 123

Qy     550 AGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGT 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     124 AGCTTGCTAGCCAGAAATGACTGTGATTCTGCTGGGGGCTGCTGAGAAGGTAATGTAGGT 183

Qy     610 TGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATTGACACTTCCACACACTCCCTGTA 669
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     184 TGAAAAGGGGCTCTAAGTTTATTTATTTCGTTAGATGGACACTTCCACACACTCCCTGTA 243

Qy     670 GTCCAGGTAGGGCCTAGAAATAGGAAAGGCTAGGATTGGATAATGCTGCAAATGCTTTTT 729
           ||||||||||||| |||||||||||||||||||| ||||||||||||||||||||||||
Db     244 GTCCAGGTAGGGCCCAGAAATAGGAAAGGCTAGGATGGGATAATGCTGCAAATGCTTTTT 303

Qy     730 TTGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTTGGGAGAGAATGTCCCAAAATTGA 789
           | |||||||||||||||||||||||||||||||| |||||||||||||||||||| ||
Db     304 TGGTGTGAGAAACTGGAGAGATGTGATTTCTCCTTTGGGGAGAGAATGTCCCAAAATGGA 363

Qy     790 TTAGGCTGAGCCTTGGGAATAGTTTGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAG 849
           |||||||||||||| |||||||||| |||||||||||||||||||||||||||||||||
Db     364 TTAGGCTGAGCCTGGGGAATAGTTGGGCAGGTTTAACATCCCAAGGCTAACCTAACGTAG 423

Qy     850 TTGGGAAAGGTAGATTGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAG 909
           | ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     424 TGGGGAAAGGTAGATGGAATGAGACATGTTTTCTGTGCTTCTAAGTGTTCTGTCCCTTAG 483

Qy     910 GCTGCTATTGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAAGAAAAATT 969
           |||||||| |||||||||||||||||||||||||||||||||||||| |||||||
Db     484 GCTGCTATAGCTTCATGTTTCCATTATGGCAGGTTTAGAGAATCCTTAAAAGGAAAAATG 543

Qy     970 GACTTGCTTGCCTAAAACTACAGTGCCCCCTT 1001
           |||||||||||||||| ||||||  ||||||
Db     544 GACTTGCTTGCCTAAAAGTACAGTTGCCCCTT 575


Search completed: September 12, 2002, 16:05:23
```

Job time: 22122 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:44:25 ; Search time 28.24 Seconds
(without alignments)
411.714 Million cell updates/sec

Title: US-09-824-178-4
Perfect score: 625
Sequence: 1 MSQKQEEENPAEETQDEKQD.....DLPQRKSSLVTSKLAGGQVE 121

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	425.5	68.1	112	2	B35308	cAMP-regulated pho
2	406	65.0	77	2	JC4878	alpha endosulfine
3	128.5	20.6	174	2	T23557	hypothetical prote
4	122	19.5	139	2	T37510	hypothetical prote
5	95.5	15.3	842	2	G90576	heat shock atp-dep
6	91.5	14.6	2664	2	T28626	variant-specific s
7	90.5	14.5	168	2	S60970	hypothetical prote
8	85	13.6	775	2	A84526	hypothetical prote
9	83.5	13.4	407	1	EDBEQ3	immediate-early pr
10	83	13.3	619	2	A41971	surface protein ps
11	83	13.3	619	2	A97887	surface protein ps
12	83	13.3	763	2	T08929	hypothetical prote
13	83	13.3	792	2	E84525	hypothetical prote
14	82.5	13.2	795	2	D84477	hypothetical prote
15	82	13.1	335	2	T33457	hypothetical prote
16	82	13.1	644	2	S55395	neurofilament prot
17	81.5	13.0	473	2	S25776	transforming prote
18	81.5	13.0	646	2	T02398	hypothetical prote
19	81.5	13.0	751	2	S65469	DNA topoisomerase
20	81.5	13.0	789	2	E84514	hypothetical prote
21	81	13.0	155	2	JQ0797	histone H2B.IV - v
22	81	13.0	301	2	T33068	hypothetical prote
23	81	13.0	325	2	I38616	zinc finger protei
24	81	13.0	695	2	I51652	dsRNA-binding prot
25	80.5	12.9	522	2	C96608	hypothetical prote
26	80.5	12.9	820	2	T12972	hypothetical prote
27	80	12.8	215	2	T23482	hypothetical prote
28	79	12.6	785	2	C88131	protein F10G7.1 [i
29	79	12.6	858	2	S15762	neurofilament trip
30	79	12.6	1014	2	T31433	Na+/Ca2+,K+-exchan
31	79	12.6	4967	2	S72269	ryanodine receptor
32	78.5	12.6	260	2	S60480	low temperature-in
33	78.5	12.6	1359	2	T34036	hypothetical prote
34	78	12.5	411	2	S48647	peptidylprolyl iso
35	78	12.5	535	2	T41384	hypothetical prote
36	78	12.5	649	2	T37740	coiled-coil protei
37	78	12.5	742	2	T48930	hypothetical prote
38	78	12.5	792	2	H83736	subtilisin-type al
39	77.5	12.4	798	2	I50479	neurofilament medi
40	77.5	12.4	806	2	T23648	hypothetical prote
41	77.5	12.4	1090	2	H86806	hypothetical prote
42	77.5	12.4	1351	2	C71607	hypothetical prote
43	77.5	12.4	2109	2	E89066	protein H05O09.1 [
44	77.5	12.4	2109	2	T33247	hypothetical prote
45	77	12.3	366	2	AC1261	hypothetical prote

ALIGNMENTS

RESULT 1
B35308
cAMP-regulated phosphoprotein, 19K - bovine
N;Alternate names: ARPP-19
N;Contains: cAMP-regulated phosphoprotein, 16K
C;Species: Bos primigenius taurus (cattle)
C;Date: 17-Aug-1990 #sequence_revision 17-Aug-1990 #text_change 24-Nov-1999
C;Accession: B35308; A35308
R;Horiuchi, A.; Williams, K.R.; Kurihara, T.; Nairn, A.C.; Greengard, P.
J. Biol. Chem. 265, 9476-9484, 1990
A;Title: Purification and cDNA cloning of ARPP-16, a cAMP-regulated phosphoprotein en
A;Reference number: A35308; MUID:90264450
A;Accession: B35308
A;Molecule type: mRNA
A;Residues: 1-112 <HOR>
A;Cross-references: GB:M33618; NID:g162689; PIDN:AAA30386.1; PID:g162690; GB:J05450
A;Accession: A35308
A;Molecule type: mRNA
A;Residues: 17-112 <HO2>
A;Cross-references: GB:M33617; NID:g162687; PIDN:AAA30385.1; PID:g162688; GB:J05450
C;Keywords: acetylated amino end; phosphoprotein
F;2-112/Product: cAMP-regulated phosphoprotein, 19K #status predicted <MAT>
F;17-112/Product: cAMP-regulated phosphoprotein, 16K #status predicted <MA2>
F;2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental
F;17/Modified site: acetylated amino end (Met) (in mature form) #status predicted

Query Match 68.1%; Score 425.5; DB 2; Length 112;
Best Local Similarity 76.4%; Pred. No. 1.9e-29;
Matches 84; Conservative 9; Mismatches 16; Indels 1; Gaps 1;

```
Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDS 67
           | |   : ||::: ::|    ||:||||||||:|| |||||||||| ||||||||||
Db       4 EVPEAASAEEQKEMEDKV-TSPEKAEEAKLKARYPHLGQKPGGSDFLRKRLQKGQKYFDS 62

Qy      68 GDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           ||||||||||||||||:| |||  |||||||||||||||| ||| ||||
Db      63 GDYNMAKAKMKNKQLPTATPDKTEVTGDHIPTPQDLPQRKPSLVASKLAG 112
```

RESULT 2
JC4878
alpha endosulfine - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 30-Aug-1996 #sequence_revision 18-Oct-1996 #text_change 05-Nov-1999
C;Accession: JC4878
R;Peyrollier, K.; Heron, L.; Virsolvy-Vergine, A.; Le Cam, A.; Bataille, D.
Biochem. Biophys. Res. Commun. 223, 583-586, 1996
A;Title: Alpha endosulfine is a novel molecule, structurally related to a family of p
A;Reference number: JC4878; MUID:96268722
A;Contents: brain
A;Accession: JC4878

```
A;Molecule type: mRNA
A;Residues: 1-77 <PEY>
A;Cross-references: EMBL:X95958; NID:g1502303; PIDN:CAA65196.1; PID:e224652; PID:g150230
A;Experimental source: brain
C;Comment: This protein structurally related to a phosphoprotein family.

  Query Match             65.0%;  Score 406;  DB 2;  Length 77;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-28;
  Matches   77;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      25 EGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPS 84
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 EGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPS 60

Qy      85 AGPDKNLVTGDHIPTPQ 101
           |||||||||||||||||
Db      61 AGPDKNLVTGDHIPTPQ 77


RESULT   3
T23557
hypothetical protein K10C3.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T23557
R;Mortimore, B.
submitted to the EMBL Data Library, December 1996
A;Reference number: Z19759
A;Accession: T23557
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-174 <WIL>
A;Cross-references: EMBL:Z83235; PIDN:CAB05775.1; GSPDB:GN00019; CESP:K10C3.2
A;Experimental source: clone K10C3
C;Genetics:
A;Gene: CESP:K10C3.2
A;Map position: 1
A;Introns: 56/3; 132/3; 162/3

  Query Match             20.6%;  Score 128.5;  DB 2;  Length 174;
  Best Local Similarity   29.0%;  Pred. No. 0.00055;
  Matches   38;  Conservative   16;  Mismatches   24;  Indels   53;  Gaps    5;

Qy      29 PERAEEAKLKAKYPSLGQKPG--GSDFLMKRLQKGQKYFDSGDYNMAKAK----MKNKQL 82
           ||: :| :|  |  : |: |    | || |:|: :|:|||||| | |:|    : :|
Db      22 PEKQQEQELMGKLAATGKLPARPASSFLQKKLQQ-RKFFDSGDYAMDKSKAGTGLGSKPH 80

Qy      83 PSAG-------------------------------------------PDKNLVTGDHIPTP 100
           | ||                                           | ||     || |
Db      81 PLAGGPPPAAPPVVAQRSPAPAATTPSPSASPISQQTNRPSSDRNSDDDNL----QIPRP 136

Qy     101 QDLPQRKSSLV 111
           :|||||:|::
Db     137 DTVPQRKASII 147


RESULT   4
T37510
hypothetical protein SPAC10F6.16 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T37510
R;Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, November 1997
A;Reference number: Z21719
A;Accession: T37510
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-139 <CON>
A;Cross-references: EMBL:AL009197; PIDN:CAA15729.1; GSPDB:GN00066; SPDB:SPAC10F6.16
A;Experimental source: strain 972h-; cosmid c10F6
C;Genetics:
A;Gene: SPDB:SPAC10F6.16
A;Map position: 1
A;Introns: 28/3; 56/3

  Query Match             19.5%;  Score 122;  DB 2;  Length 139;
  Best Local Similarity   34.1%;  Pred. No. 0.0015;
  Matches   30;  Conservative   15;  Mismatches   31;  Indels   12;  Gaps    3;

Qy      20 DTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKN 79
           ::::|  :     || ||   |   |: |   | |:::||:|:|||||||| : ||
Db      20 NSEQKVDVAKLSPEEQKLFRLY---GRLPQRKDLLVQKLQQGRKYFDSGDYALNKA---- 72

Qy      80 KQLPSAGPDKNLV-TGDHIPTPQDLPQR 106
                  |  :   |  ||:|  :| |
Db      73 ----GKASDSGITCIGKEIPSPDTIPHR 96


RESULT   5
G90576
heat shock atp-dependent proteinase [imported] - Mycoplasma pulmonis (strain UAB CTIP
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: G90576
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer,
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: G90576
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-842 <KUR>
A;Cross-references: GB:AL445566; PID:g14089933; PIDN:CAC13692.1; GSPDB:GN00153
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU_5190
A;Genetic code: SGC3
C;Superfamily: ATP-dependent serine proteinase La

  Query Match             15.3%;  Score 95.5;  DB 2;  Length 842;
  Best Local Similarity   31.5%;  Pred. No. 2;
  Matches   35;  Conservative   19;  Mismatches   48;  Indels    9;  Gaps    4;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQ-KPGGSDFLMKRLQKG 61
           : |: ||  :|  ::    |:|| || |     |:||    ||:      ::  |   |
Db     234 EDQDLENAIDEKMKKNLSKQQKEFILRE-----KMKAIRSELGESNDDEEEYKTKIKDKN 288

Qy      62 QK--YFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSL 110
           |  |  |    :|  : | | : |: || | :| ::| |   || ||:|:
Db     289 NKKIYPKSVRKLIANEESKLKNMMSSSPDAN-ITRNYIDTLMKLPWRKTSI 338


RESULT   6
T28626
variant-specific surface protein 2 - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T28626
R;Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herrfeldt, J.A.; Peterson,
Cell 82, 89-100, 1995
A;Title: The large diverse gene family var encodes proteins involved in cytoadherence
A;Reference number: Z20487; MUID:95330813
A;Accession: T28626
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2664 <SUX>
A;Cross-references: EMBL:L40609; NID:g886376; PID:g886378; PIDN:AAA75398.1
C;Genetics:
A;Introns: 2197/3
A;Note: var-2
```

```
  Query Match             14.6%;  Score 91.5;  DB 2;  Length 2664;
  Best Local Similarity   26.8%;  Pred. No. 15;
  Matches   34;  Conservative   16;  Mismatches   30;  Indels   47;  Gaps    5;

Qy       6 EEENPAEETG-----EEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           |||||  |: |     ::: :  :| | | |: :| | : : |:  :: ||:
Db    2086 EEENPVEQPGFCPTPQQEPEPDDKCGKLEEKKDEKKEQPEQPA--EEDGGA--------- 2134

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGP------DKNLVTGDHIPTPQDLPQRK-----SS 109
                              :||  |      ||  |    || ||: |          |
Db    2135 --------------------IVPSGPPGSEPEADKGPVKPAEIPKPQEPPDLSHPAVIPS 2174

Qy     110 LVTSKLA 116
           |||| ||
Db    2175 LVTSTLA 2181


RESULT   7
S60970
hypothetical protein YNL157w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein N1743
C;Species: Saccharomyces cerevisiae
C;Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jul-2000
C;Accession: S60970; S63109; S63817
R;Nasr, F.; Becam, A.M.; Herbert, C.J.
submitted to the EMBL Data Library, October 1995
A;Description: The sequence of 36.8 kb from the left arm of chromosome XIV reveals 24 co
tonic dystrophy kinase.
A;Reference number: S60958
A;Accession: S60970
A;Molecule type: DNA
A;Residues: 1-168 <NAS>
A;Cross-references: EMBL:X92517; NID:g1050783; PID:g1050795
R;Nasr, F.; Becam, A.M.; Herbert, C.
submitted to the Protein Sequence Database, April 1996
A;Reference number: S62967
A;Accession: S63109
A;Molecule type: DNA
A;Residues: 1-168 <NAW>
A;Cross-references: EMBL:Z71433; NID:g1302119; PID:e239815; PID:g1302120; MIPS:YNL157w
A;Experimental source: strain S288C
R;Nasr, F.; Becam, A.M.; Herbert, C.J.
Yeast 12, 169-175, 1996
A;Title: The sequence of 36.8 kb from the left arm of chromosome XIV reveals 24 complete
dystrophy kinase.
A;Reference number: S63805; MUID:96287653
A;Accession: S63817
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-168 <NAF>
A;Cross-references: EMBL:X92517; NID:g1050783; PIDN:CAA63282.1; PID:g1050795
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995
C;Genetics:
A;Map position: 14L
A;Note: YNL157w


  Query Match             14.5%;  Score 90.5;  DB 2;  Length 168;
  Best Local Similarity   31.9%;  Pred. No. 0.89;
  Matches   30;  Conservative   11;  Mismatches   30;  Indels   23;  Gaps    5;

Qy      34 EAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKN--------KQLPSA 85
           | ||   |   |: |   |      :|| :|||||||| : || ::|          ||
Db      35 EMKL---YKMYGKLPSKKDIFKHTMQK-RKYFDSGDYALQKAGIQNNDPINYGKNNLPLT 90

Qy      86 GPDKNLVTGDHIPTPQDLPQRKSSLV--TSKLAG 117
             | |           :|: :|: |    |:  ||
Db      91 NPSK---------LREDIIKRRISTCPSTASTAG 115


RESULT   8
A84526
hypothetical protein At2g15190 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: A84526
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: A84526
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-775 <STO>
A;Cross-references: GB:AE002093; NID:g4585912; PIDN:AAD25573.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g15190
A;Map position: 2
C;Superfamily: Arabidopsis thaliana hypothetical protein T21L8.10


  Query Match             13.6%;  Score 85;  DB 2;  Length 775;
  Best Local Similarity   20.6%;  Pred. No. 14;
  Matches   32;  Conservative   24;  Mismatches   43;  Indels   56;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAE------EAKLKAKYPSLGQKPGGSDFLMK- 56
           :::||   |: | |:|   |:|| |  :||      |  ::      |  ||  :  :
Db     375 EEDEERKQEDEGVERQPEAEEEGGLERKAENDNESFEDSIREPNTQYGSYPGDDENTQRD 434

Qy      57 --------------------------------RLQKGQKYFDSGDYNMAKAKMKNKQL-P 83
                                           :|| :|        ::: : :||:: |
Db     435 VGDEVAEESSKDKSPTPRSSTPNFNILSEESLDVQKDKK-------RVSRGRNENKRVKP 487

Qy      84 SAGPDKNLVTGDHIPTPQ---------DLPQRKSS 109
           |   : || |   :|  :          |:| || |
Db     488 SVHAEDNLKTRKQVPRKRQKQVDSADVDVPTRKES 522


RESULT   9
EDBEQ3
immediate-early protein - saimiriine herpesvirus 1 (strain 11)
C;Species: saimiriine herpesvirus 1
A;Note: host Saimiri sciureus (common squirrel monkey)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
C;Accession: G36813; S20244
R;Albrecht, J.
submitted to the EMBL Data Library, January 1992
A;Description: Primary structure of the herpesvirus saimiri genome.
A;Reference number: A36806
A;Accession: G36813
A;Molecule type: DNA
A;Residues: 1-407 <ALB>
A;Cross-references: GB:X64346; NID:g60320; PIDN:CAA45696.1; PID:g60394
R;Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.; Newman, C.;
J. Virol. 66, 5047-5058, 1992
A;Title: Primary structure of the herpesvirus saimiri genome.
A;Reference number: A37309; MUID:92333688
A;Contents: annotation; possible protein-coding frames
A;Note: neither amino acid nor nucleotide sequence is given
R;Nicholas, J.; Cameron, K.R.; Honess, R.W.
Nature 355, 362-365, 1992
A;Title: Herpesvirus saimiri encodes homologues of G protein-coupled receptors and cy
A;Reference number: S20243; MUID:92115001
A;Accession: S20244
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-407 <NIC>
A;Cross-references: GB:S76368; NID:g243351; PIDN:AAB21116.1; PID:g243353
C;Genetics:
A;Gene: 73
C;Superfamily: herpesvirus immediate-early protein IE68
C;Keywords: early protein
```

```
  Query Match             13.4%;  Score 83.5;  DB 1;  Length 407;
  Best Local Similarity   30.6%;  Pred. No. 9.3;
  Matches   33;  Conservative   13;  Mismatches   35;  Indels   27;  Gaps    5;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAK----------YPSLGQK---- 47
           : :||   |||  || :: :| |    | ||||: : :           |  :|||
Db     203 EAEEEAEEAEEEAEEAEEAEEAEEAE-AEEEAEEAEEEEEEEAGPSTPRLPHYKVVGQKPSTQ 261

Qy      48 PGGSDFL--------MKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGP 87
             |||   |          || ||::   | |    | : :||    | |
Db     262 PGGVPKLCLKMQPQHRSRLPKGKQ---SHDKVPKKYQARNKFFSQAAP 306


RESULT  10
A41971
surface protein pspA precursor - Streptococcus pneumoniae
N;Alternate names: pneumococcal surface protein A
C;Species: Streptococcus pneumoniae
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C;Accession: A41971; A60282; A33134
R;Yother, J.; Briles, D.E.
J. Bacteriol. 174, 601-609, 1992
A;Title: Structural properties and evolutionary relationships of PspA, a surface protein
A;Reference number: A41971; MUID:92105030
A;Accession: A41971
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-619 <YOT>
A;Cross-references: GB:M74122; NID:g153840; PIDN:AAA27018.1; PID:g153841
A;Note: sequence extracted from NCBI backbone (NCBIN:75635, NCBIP:75636)
R;Talkington, D.F.; Crimmins, D.L.; Voellinger, D.C.; Yother, J.; Briles, D.E.
Infect. Immun. 59, 1285-1289, 1991
A;Title: A 43-kilodalton pneumococcal surface protein, PspA: isolation, protective abili
A;Reference number: A60282; MUID:91169598
A;Accession: A60282
A;Molecule type: protein
A;Residues: 32-76 <TAL>
A;Experimental source: strain JY2008
C;Genetics:
A;Gene: pspA
C;Superfamily: cpl repeat homology
F;1-31/Domain: signal sequence #status predicted <SIG>
F;32-619/Product: surface protein pspA #status predicted <MAT>
F;411-430/Domain: cpl repeat homology <CP01>
F;431-450/Domain: cpl repeat homology <CP02>
F;451-470/Domain: cpl repeat homology <CP03>
F;471-490/Domain: cpl repeat homology <CP04>
F;491-510/Domain: cpl repeat homology <CP05>
F;511-530/Domain: cpl repeat homology <CP06>
F;531-550/Domain: cpl repeat homology <CP07>
F;551-570/Domain: cpl repeat homology <CP08>
F;571-591/Domain: cpl repeat homology <CP09>
F;592-611/Domain: cpl repeat homology <CP10>


  Query Match             13.3%;  Score 83;  DB 2;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 16;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  11
A97887
surface protein pspA precursor [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C;Accession: A97887
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: A97887
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-619 <KUR>
A;Cross-references: GB:AE007317; PIDN:AAK98925.1; PID:g15457660; GSPDB:GN00174
C;Genetics:
A;Gene: pspA


  Query Match             13.3%;  Score 83;  DB 2;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 16;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  12
T08929
hypothetical protein T15N24.80 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C;Accession: T08929
R;Bevan, M.; Zimmermann, W.; Grueneisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.;
submitted to the Protein Sequence Database, May 1999
A;Reference number: Z16518
A;Accession: T08929
A;Molecule type: DNA
A;Residues: 1-763 <BEV>
A;Cross-references: EMBL:AL078465; GSPDB:GN00062; ATSP:T15N24.80
A;Experimental source: cultivar Columbia; BAC clone T15N24
C;Genetics:
A;Gene: ATSP:T15N24.80
A;Map position: 4
A;Introns: 345/3; 357/1; 384/3; 404/3; 441/3; 472/3; 500/3; 667/1; 697/3; 708/1


  Query Match             13.3%;  Score 83;  DB 2;  Length 763;
  Best Local Similarity   25.0%;  Pred. No. 20;
  Matches   31;  Conservative   25;  Mismatches   36;  Indels   32;  Gaps    6;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGI--LPERAEEAKLKAKYPSLGQKPGGSDFLMK---- 56
           :|:|||   ||  |||:: :|:|   :|:::|:   :|  ||  ::   |:   :
Db     526 EKEEEEKQEEEKAEEKEEKKEEENENGIPDKSED---EAPQPSESEEKDESEEHSEEETT 582

Qy      57 ------RLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSL 110
                 ||  |:|         :  : :||       | :|    .| |: : |::||
Db     583 KKKRGSRLSAGKK--------ESAGRARNK--------KAVVAAKSSP-PEKITQKRSSA 625

Qy     111 VTSK 114
               |
Db     626 KRKK 629


RESULT  13
E84525
hypothetical protein At2g15140 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
```

```
C;Accession: E84525
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: E84525
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-792 <STO>
A;Cross-references: GB:AE002093; NID:g4585915; PIDN:AAD25576.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g15140
A;Map position: 2
C;Superfamily: Arabidopsis thaliana hypothetical protein T21L8.10


  Query Match            13.3%;  Score 83;  DB 2;  Length 792;
  Best Local Similarity  21.3%;  Pred. No. 21;
  Matches   33;  Conservative   25;  Mismatches   41;  Indels   56;  Gaps    6;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAE------EAKLKAKYPSLGQKPGG------- 50
           :::||   |: | |:|   |:|| |  :||      |  ::       |  ||
Db     386 EEDEERKQEDEGVERQPEAEEEGGLERKAENDNESFEDSIREPNTQYGSYPGDDENTQRD 445

Qy      51 -SDFLMKR-------------------------LQKGQKYFDSGDYNMAKAKMKNKQL-P 83
             | |::                          :|| :|        ::: : :||:: |
Db     446 VGDELVEESSKDKSPTPRSSTPNFNILSEESLDVQKDKK-------RVSRGRNQNKRVKP 498

Qy      84 SAGPDKNLVTGDHIPTPQ---------DLPQRKSS 109
           |   : || |   :|  :         |:| || :
Db     499 SVHAEDNLKTRKQVPRKRQKQVDSADVDVPTRKEA 533


RESULT  14
D84477
hypothetical protein At2g06430 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: D84477
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: D84477
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-795 <STO>
A;Cross-references: GB:AE002093; NID:g6598730; PIDN:AAD26891.2; GSPDB:GN00139
C;Genetics:
A;Gene: At2g06430
A;Map position: 2
C;Superfamily: Arabidopsis thaliana hypothetical protein T21L8.10


  Query Match            13.2%;  Score 82.5;  DB 2;  Length 795;
  Best Local Similarity  25.4%;  Pred. No. 23;
  Matches   33;  Conservative   22;  Mismatches   38;  Indels   37;  Gaps    6;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKR---LQK 60
           :::||   |: | |:|   |:|| |  :||:     | |:        : | :    :||
Db     405 EEDEERKQEDEGVERQPEAEEEGGLERKAED-----KSPTPRSTTPNFNILSEESLDVQK 459

Qy      61 GQKYFDSGDYNMAKAKMKNKQL-PSAGPDKNLVT--------------GDHIPTP---- 100
             :|        ::: : :||:: ||   : || |               ||:
Db     460 DKK-------RVSRGRNENKRVKPSVHAEDNLKTRKQAQSKKRKIIGNDGDNADNDGDND 512

Qy     101 --QDLPQRKS 108
             |  |||||
Db     513 DFQPAPQRKS 522


RESULT  15
T33457
hypothetical protein F36H12.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33457
R;Blanchard, M.; Bradshaw, H.; Stellyes, L.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of C. elegans cosmid F36H12.
A;Reference number: Z21346
A;Accession: T33457
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-335 <BLA>
A;Cross-references: EMBL:AF078790; PIDN:AAC26930.1; GSPDB:GN00022; CESP:F36H12.3
A;Experimental source: strain Bristol N2; clone F36H12
C;Genetics:
A;Gene: CESP:F36H12.3
A;Map position: 4
A;Introns: 32/3; 227/1; 270/2


  Query Match            13.1%;  Score 82;  DB 2;  Length 335;
  Best Local Similarity  27.2%;  Pred. No. 10;
  Matches   41;  Conservative   20;  Mismatches   56;  Indels   34;  Gaps    7;

Qy       3 QKQEE--ENPAEETGEEKQDTQEKEGILPERAE-EAKLKAK----YPSLGQKPGGSDFLM 55
           :|:||  |   ||  |||:: ::||    |: | | |:: |     | :  :| | | :
Db     179 EKKEEKKEEKKEEKKEEKKEEEKKEEKKEEKKEDEKKVEPKPGEPKPHISIEPPG-DLMF 237

Qy      56 KRLQKGQKYFDSGDYN----MAKAKMKNKQLPSAGPDKNLV----TGDHIPTPQDLPQRK 107
           |  |: ||     : :    | | |  : |:    |    |    | |   |    | ::
Db     238 KADQQEQKKLTLTNTHDKKIMFKIKTSDNQVYLVNPVYGTVEPGKTADLTITRNKAPAKE 297

Qy     108 SSLV-----------------TSKLAGGQV 120
           : ||                   |  ||||
Db     298 AKLVIVNSVFSGDDKDLAKSFKIGKPTGGQV 328


RESULT  16
S55395
neurofilament protein M - rabbit (fragment)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 13-Aug-1999
C;Accession: S55395
R;Vitadello, M.; Vettore, S.; Lamar, E.; Chien, K.R.; Gorza, L.
submitted to the EMBL Data Library, January 1995
A;Description: Neurofilament mRNA and protein are expressed in precursors of heart co
A;Reference number: S55395
A;Accession: S55395
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-644 <VIT>
A;Cross-references: EMBL:Z47378; NID:g854352; PIDN:CAA87454.1; PID:g854353
C;Superfamily: cytoskeletal keratin


  Query Match            13.1%;  Score 82;  DB 2;  Length 644;
  Best Local Similarity  28.9%;  Pred. No. 21;
  Matches   22;  Conservative   17;  Mismatches   27;  Indels   10;  Gaps    2;

Qy       5 QEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKY 64
           :|||   || |:|::: :| ||:  ::|||          | :  ||       ::|:
Db     308 KEEEGEKEEEGQEEEEEEEDEGVKSDQAEEG---------GSEKEGSSKNEGEQEEGETE 358

Qy      65 FDSGDYNMAKAKMKNK 80
             : |:   |:|| : |
Db     359 AE-GEVEEAEAKEEKK 373


RESULT  17
```

```
S25776
transforming protein (SHC) - human
C;Species: Homo sapiens (man)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C;Accession: S25776
R;Pelicci, G.; Lanfrancone, L.; Grignani, F.; McGlade, J.; Cavallo, F.; Forni, G.; Nicol
Cell 70, 93-104, 1992
A;Title: A novel transforming protein (SHC) with an SH2 domain is implicated in mitogeni
A;Reference number: S25776; MUID:92323554
A;Accession: S25776
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-473 <PEL>
A;Cross-references: EMBL:X68148; NID:g36453; PIDN:CAA48251.1; PID:g36454
C;Superfamily: SH2 homology
F;378-467/Domain: SH2 homology <SH2B>

  Query Match             13.0%;  Score 81.5;  DB 2;  Length 473;
  Best Local Similarity   26.2%;  Pred. No. 16;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   33;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|       | : :||  |  |      |: ||       :||
Db     229 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 288

Qy      47 KPGGSDFLMKRLQ-----KGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDHI 97
             ||   : |::       |:: ||   |    |: ||:   : :  |||    : |
Db     289 PVGGDPEVRKQMPPPPPCPGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS-- 343

Qy      98 PTPQDL 103
             |:||
Db     344 -APRDL 348


RESULT  18
T02398
hypothetical protein At2g44430 [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein F4I1.24
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C;Accession: T02398; D84878
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, May 1998
A;Description: Arabidopsis thaliana chromosome II BAC F4I1 genomic sequence.
A;Reference number: Z14667
A;Accession: T02398
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-646 <ROU>
A;Cross-references: EMBL:AC004521; NID:g3128166; PID:g3128185
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: D84878
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-646 <STO>
A;Cross-references: GB:AE002093; NID:g3128185; PIDN:AAC16089.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g44430; F4I1.24
A;Map position: 2
A;Introns: 156/2; 335/3

  Query Match             13.0%;  Score 81.5;  DB 2;  Length 646;
  Best Local Similarity   25.8%;  Pred. No. 23;
  Matches   32;  Conservative   17;  Mismatches   52;  Indels   23;  Gaps    4;

Qy       3 QKQ-EEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKG 61
           ||| | :  :      ||||              | : |  |  :|   :||| |||:|
Db     515 QKQTESKTNSSNDNSSKQDT------------GKTEKKTVSADKKKSVADFL-KRLKKN 560

Qy      62 QKYFDSGDYNM--------AKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTS 113
             :: | |          :| | :  :  | |  |  |    :      ||:|::  |:
Db     561 SPQKEAKDQNKSGGNVKKDSKTKPRELRSSSVGKKKAEVENTPVKRAPGRPQKKTAEATA 620

Qy     114 KLAG 117
             :|
Db     621 SASG 624


RESULT  19
S65469
DNA topoisomerase (EC 5.99.1.2) - Caenorhabditis elegans
N;Alternate names: DNA topoisomerase I
C;Species: Caenorhabditis elegans
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-May-1997
C;Accession: S65469
R;Kim, J.; Kim, Y.C.; Lee, J.H.; Jang, Y.J.; Chung, I.K.; Koo, H.S.
Eur. J. Biochem. 237, 367-372, 1996
A;Title: cDNA cloning, expression, and chromosomal localization of Caenorhabditis ele
A;Reference number: S65469; MUID:96215431
A;Accession: S65469
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-751 <KIM>
A;Cross-references: EMBL:X96762
C;Genetics:
A;Gene: topoI
C;Superfamily: eukaryotic type I DNA topoisomerase
C;Keywords: isomerase

  Query Match             13.0%;  Score 81.5;  DB 2;  Length 751;
  Best Local Similarity   30.1%;  Pred. No. 27;
  Matches   37;  Conservative   14;  Mismatches   31;  Indels   41;  Gaps    6;

Qy       6 EEENPAEET-GEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKY 64
           |||: |  |  ||:|  :|||    |: |: | | |                |||:|  :
Db      86 EEEDEAPTTDDEEEQKRKEKERRKKEKREKKKRKEK--------------KRLEKENRK 130

Qy      65 F------DSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAGG 118
             || | :  ||| | ::  | | :|:                 | |  : | |||
Db     131 IKEEDDEDSDDEDDEKAKKKKRK--SKGAEKS---------------KPS-TSKKDAGG 171

Qy     119 QVE 121
             : |
Db     172 KKE 174


RESULT  20
E84514
hypothetical protein At2g14130 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: E84514
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: E84514
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-789 <STO>
A;Cross-references: GB:AE002093; NID:g4587682; PIDN:AAD25852.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g14130
A;Map position: 2
C;Superfamily: Arabidopsis thaliana hypothetical protein T21L8.10
```

```
  Query Match             13.0%;  Score 81.5;  DB 2;  Length 789;
  Best Local Similarity   19.3%;  Pred. No. 28;
  Matches   28;  Conservative   28;  Mismatches   50;  Indels   39;  Gaps    4;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAE------EAKLKAKYPSLGQKPGGSDFLMK- 56
           :::||   |: | |:|   |:|| |  :||      |  ::      |  ||  :   :
Db     386 EEDEERKQEDEGVERQPEAEEEGGLERKAENDNESFEDSIREPNTQYGSYPGDDENTQRD 445

Qy      57 ----------------------RLQKGQKYFDSGDYNMAKAKMKNKQL-PSAGPDKNLVT 93
                                  : ::     : :::  :::||:: ||   : || |
Db     446 VGDEVAEESSKDKSHTPRSSTPNFNRIRRELAEEEMSLSFMQLQNKRVKPSVHAEDNLKT 505

Qy      94 GDHIPTPQ---------DLPQRKSS 109
              :|  :         |:| || :
Db     506 RKQVPRKRQKQVDSADVDVPTRKEA 530


RESULT  21
JQ0797
histone H2B.IV - Volvox carteri
C;Species: Volvox carteri
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 23-Jul-1999
C;Accession: JQ0797
R;Mueller, K.; Lindauer, A.; Bruederlein, M.; Schmitt, R.
Gene 93, 167-175, 1990
A;Title: Organization and transcription of Volvox histone-encoding genes: similarities b
A;Reference number: JQ0794; MUID:91033024
A;Accession: JQ0797
A;Molecule type: DNA
A;Residues: 1-155 <MUE>
A;Cross-references: GB:M31922; NID:g170657; PIDN:AAA34250.1; PID:g170659
C;Genetics:
A;Gene: H2B-IV
C;Superfamily: histone H2B
C;Keywords: chromosomal protein; DNA binding; nucleosome core; nucleus


  Query Match             13.0%;  Score 81;  DB 2;  Length 155;
  Best Local Similarity   29.2%;  Pred. No. 5.3;
  Matches   26;  Conservative   13;  Mismatches   42;  Indels    8;  Gaps    2;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEE---AKLKAKYPSLGQKPGGSDFLMKR 57
           |: | :|| || |  | | : : |    |:: |:   ||  || |: |    |      :
Db       1 MAPKTKEEKPASEAVEPKAEAKPKAEKAPKKKEKKAPAKKSAKEPAAGDAAEGD-----K 55

Qy      58 LQKGQKYFDSGDYNMAKAKMKNKQLPSAG 86
              :|  |   |  | :   |:  :  |  |
Db      56 KKKKAKVAKSETYKLYIYKVLKQVHPDTG 84


RESULT  22
T33068
hypothetical protein C35E7.9 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33068
R;Graves, T.; McDonald, R.
submitted to the EMBL Data Library, May 1998
A;Description: The sequence of C. elegans cosmid C35E7.
A;Reference number: Z21278
A;Accession: T33068
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-301 <GRA>
A;Cross-references: EMBL:AF067216; PIDN:AAC17524.1; GSPDB:GN00019; CESP:C35E7.9
A;Experimental source: strain Bristol N2; clone C35E7
C;Genetics:
A;Gene: CESP:C35E7.9
A;Map position: 1
A;Introns: 30/3; 193/1; 236/2


  Query Match             13.0%;  Score 81;  DB 2;  Length 301;
  Best Local Similarity   26.4%;  Pred. No. 11;
  Matches   39;  Conservative   18;  Mismatches   51;  Indels   40;  Gaps    6;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQ 62
           :|:|||   |:  |:|:: :| |    |: || |     | :   | | | : |  |: |
Db     157 EKKEEEKKEEKKEEKKEEKKEDEKKEDEKKEEPK-----PHITIDPPG-DLMFKADQQEQ 210

Qy      63 KYFDSGDYN----MAKAKMKNKQLPSAGPDKNLVTGDHIP--------TPQDLPQRKSSL 110
           |     : :    | | |  : |:    |    | |   |         |    | ::: |
Db     211 KKLKLTNTHDKKIMFKIKTSDNQVYLMNP----VYGTVEPGKSANLTLTRNKAPAKEAKL 266

Qy     111 V------------------TSKLAGGQV 120
           |                  | |  ||||
Db     267 VIVNSVFSGDDKDLAKSFKTGKPTGGQV 294


RESULT  23
I38616
zinc finger protein ZNF139 - human (fragment)
N;Alternate names: finger protein kox18
C;Species: Homo sapiens (man)
C;Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 01-Dec-2000
C;Accession: I38616; I37957; S10406
R;Tommerup, N.; Vissing, H.
Genomics 27, 259-264, 1995
A;Title: Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs iden
A;Reference number: A57785; MUID:96044430
A;Accession: I38616
A;Molecule type: mRNA
A;Residues: 1-325 <RES>
A;Cross-references: EMBL:U09848; NID:g495567; PIDN:AAC50264.1; PID:g495568
R;Thiesen, H.J.
New Biol. 2, 363-374, 1990
A;Title: Multiple genes encoding zinc finger domains are expressed in human T cells.
A;Reference number: I37949; MUID:91145339
A;Accession: I37957
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 167-222 <RE2>
A;Cross-references: EMBL:X52349; NID:g34150; PIDN:CAA36575.1; PID:g930080
C;Genetics:
A;Gene: GDB:ZNF139
A;Cross-references: GDB:137044
A;Map position: 7q21.3-7q22.1
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: DNA binding; zinc finger


  Query Match             13.0%;  Score 81;  DB 2;  Length 325;
  Best Local Similarity   21.3%;  Pred. No. 12;
  Matches   32;  Conservative   25;  Mismatches   47;  Indels   46;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           :||    |   | | ::  |:     | :|::  : :     ||    :|  || |:|:
Db      18 RQENYGSAFPQGGENRNENEESTSKAETSEDSASRGETTGRSQK----EFGEKRDQEGK- 72

Qy      64 YFDSGDYNMAKAKMKNKQL-----PSAGPDKNLVTGDHIP-----------------TP 100
             :|:      : | ::      |: | ||  :||:  |                   ||
Db      73 ---TGERQQKNPEEKTRKEKRDSGPAIGKDKKTITGERGPREKGKGLGRSFSLSSNFTTP 129

Qy     101 QDLP--------------QRKSSLVTSKL 115
           :::|               | |||:  |:
Db     130 EEVPTGTKSHRCDECGKCFTRSSSLIRHKI 159


RESULT  24
I51652
dsRNA-binding protein 4F.1 - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: I51652
```

```
R;Bass, B.L.; Hurst, S.R.; Singer, J.D.
Curr. Biol. 4, 301-314, 1994
A;Title: Binding properties of newly identified Xenopus proteins containing dsRNA-bindin
A;Reference number: I51652; MUID:95006648
A;Accession: I51652
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-695 <BAS>
A;Cross-references: EMBL:U07155; NID:g463902; PIDN:AAA19960.1; PID:g514259
C;Superfamily: double-stranded RNA-binding repeat homology
F;401-469/Domain: double-stranded RNA-binding repeat homology <DSR1>


  Query Match             13.0%;  Score 81;  DB 2;  Length 695;
  Best Local Similarity   21.6%;  Pred. No. 27;
  Matches   36;  Conservative   19;  Mismatches   48;  Indels   64;  Gaps    7;

Qy       1 MSQKQE-----EENPAEETGEEKQDTQEKEGILPERAEEAKLKA---------------- 39
           | :|:|     |: |  :|  :  |: | :    ::|| || :
Db     472 MEEKEEGTDESEQKPVVQTPAQPDDSAEVDSAALDQAESAKQQGPILTKHGKNPVMELNE 531

Qy      40 -----KYPSLGQKPGGSD--FLMKRLQKGQKYFDSG--------------------DYNM 72
              ||  : :  |  |  |:|:    | |:  ||                      || |
Db     532 KRRGLKYEMISETGGSHDKRFVMEVEVDGVKFQGSGSNKKVAKAYAALSALEKLFPDYTM 591

Qy      73 ---AKAKMKNKQLPSAGPDKNLVTGDH----------IPTPQDLPQR 106
             |  | :   :|  ||      | |          :| || :  |
Db     592 YTEAPKKKRPPMMPRGGPK---FAGKHNQGFGMMYSEVPPPQAMRGR 635


RESULT  25
C96608
hypothetical protein F25P12.91 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: C96608
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: C96608
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-522 <STO>
A;Cross-references: GB:AE005173; NID:g9954748; PIDN:AAG09099.1; GSPDB:GN00141
C;Genetics:
A;Gene: F25P12.91
A;Map position: 1


  Query Match             12.9%;  Score 80.5;  DB 2;  Length 522;
  Best Local Similarity   19.7%;  Pred. No. 22;
  Matches   25;  Conservative   27;  Mismatches   56;  Indels   19;  Gaps    2;

Qy       3 QKQEEENPAEETGEEKQDTQEKEG------ILPERAEEAKLKAKYPSLGQKPGGSDFLMK 56
           :| : :   | :|: |:| ::|:|      :  :: :: |   |  |   : |  |  :|
Db      43 EKVKAKKDEESSGKSKKDKEKKKGKNVDSEVKEDKDDDKKKDGKMVSKKHEEGHGDLEVK 102

Qy      57 ------------RLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDL 103
                        :  | :|: :  :    | |   |:   :||::     |     :|:
Db     103 ESDVKVEEHEKEHKKGKEKKHEELEEEKEGKKKKNKKEKDESGPEEKNKKADKEKKHEDV 162

Qy     104 PQRKSSL 110
             | |  |
Db     163 SQEKEEL 169


RESULT  26
T12972
hypothetical protein T21L8.10 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 18-Aug-2000
C;Accession: T12972
R;Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.;
submitted to the Protein Sequence Database, July 1999
A;Reference number: Z17586
A;Accession: T12972
A;Molecule type: DNA
A;Residues: 1-820 <CHO>
A;Cross-references: EMBL:AL096860; GSPDB:GN00061; ATSP:T21L8.10
A;Experimental source: cultivar Columbia; BAC clone T21L8
C;Genetics:
A;Gene: ATSP:T21L8.10
A;Map position: 3
A;Introns: 115/1; 297/3; 431/3; 443/3; 484/3; 520/3; 539/3; 560/3; 641/2; 668/3; 726/
C;Superfamily: Arabidopsis thaliana hypothetical protein T21L8.10


  Query Match             12.9%;  Score 80.5;  DB 2;  Length 820;
  Best Local Similarity   20.1%;  Pred. No. 36;
  Matches   28;  Conservative   28;  Mismatches   52;  Indels   31;  Gaps    4;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAE------EAKLKAKYPSLGQKPGGSDFLMKR 57
           :::||   |: | |:|   |:|| |  :||      |  ::      |  ||  : : :
Db     414 EEDEERKQEDEGVERQPEAEEEGGLERKAENDNESFEDSIREPNTQYGSYPGDDENIQRD 473

Qy      58 LQKGQKYFDSGDYNMAKAK----------------MKNKQLPSAGPDKN------LVTG 94
           : | :  :    :|:                     |:|:  | | ::|      :
Db     474 V--GDELVEESSKDMSPTPRSSTLNFNILSEESLDVQKDKKRVSRGRNENKRVKPSVHAE 531

Qy      95 DHIPTPQDLPQRKSSLVTS 113
           |:  | : :|:::   | |
Db     532 DNFKTRKQMPRKRQKQVDS 550


RESULT  27
T23482
hypothetical protein K08F4.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T23482
R;Hembry, C.
submitted to the EMBL Data Library, January 1996
A;Reference number: Z19746
A;Accession: T23482
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-215 <WIL>
A;Cross-references: EMBL:Z68879; PIDN:CAA93085.1; GSPDB:GN00022; CESP:K08F4.5
A;Experimental source: clone K08F4
C;Genetics:
A;Gene: CESP:K08F4.5
A;Map position: 4
A;Introns: 22/1


  Query Match             12.8%;  Score 80;  DB 2;  Length 215;
  Best Local Similarity   27.2%;  Pred. No. 9.1;
  Matches   34;  Conservative   21;  Mismatches   40;  Indels   30;  Gaps    6;

Qy       3 QKQEEENPA---EETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQ 59
           :|:||:  |   ||  |||:: : ||    |: ||:|   |  |  :|    |   :: :
Db      79 KKEEEKKDATKKEEAKEEKKEEERKE----EKKEESKEGEKKNSGEEKKNSGD---EKKE 131

Qy      60 KGQKYFDSGD-YNMAKAKMKNKQLPSAGPDKNLV-----------TGDHI-------PT 99
           :| |  |: |     :||  :|:     | | :|              | :       ||
Db     132 EGGKKDDAKDGEKKEEAKEGDKKDEDQKPKKEIVPEVFKRTSKEKAIADGVKKKKTDYPT 191

Qy     100 PQDLP 104
```

```
          :|:|
Db    192 MEDVP 196


RESULT  28
C88131
protein F10G7.1 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: C88131
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: C88131
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-785 <STO>
A;Cross-references: GB:chr_II; PIDN:AA81131.1; PID:g1055162; GSPDB:GN00020; CESP:F10G7.1
C;Genetics:
A;Gene: F10G7.1
A;Map position: 2


  Query Match             12.6%;  Score 79;  DB 2;  Length 785;
  Best Local Similarity   23.1%;  Pred. No. 46;
  Matches   24;  Conservative   16;  Mismatches   44;  Indels   20;  Gaps    3;

Qy      6 EEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
          :|::  ||  :|  |  : | :  :  ||  :       |:               :  |
Db    378 DEDDEDEEDSDEDMDDSDNEEVEDDSEEEEPMDDLKSEAGETTA----------SEMMF 426

Qy     66 DSG---DYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQR 106
          | |   | |||: :   |:       :|    | : || |:| |
Db    427 DDGIDEDINMAEVEKYRKE------RENAQWPDEVDTPMDMPAR 464


RESULT  29
S15762
neurofilament triplet M protein - chicken
C;Species: Gallus gallus (chicken)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 13-Aug-1999
C;Accession: S15762; S08061; A27040
R;Zopf, D.; Dineva, B.; Betz, H.; Gundelfinger, E.D.
Nucleic Acids Res. 18, 521-529, 1990
A;Title: Isolation of the chicken middle-molecular weight neurofilament (NF-M) gene and
A;Reference number: S15762; MUID:90174973
A;Accession: S15762
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-858 <ZOP>
A;Cross-references: EMBL:X17102; NID:g63688; PIDN:CAA34958.1; PID:g63689
R;Zopf, D.; Dineva, B.; Betz, H.; Gundelfinger, E.
submitted to the EMBL Data Library, November 1989
A;Reference number: S08061
A;Accession: S08061
A;Molecule type: DNA
A;Residues: 1-355,368-858 <ZO2>
A;Cross-references: EMBL:X17102
R;Zopf, D.; Hermans-Borgmeyer, I.; Gundelfinger, E.D.; Betz, H.
Genes Dev. 1, 699-708, 1987
A;Title: Identification of gene products expressed in the developing chick visual system
A;Reference number: A27040; MUID:88112814
A;Accession: A27040
A;Molecule type: mRNA
A;Residues: 350-546,'R',548-858 <ZO3>
A;Cross-references: GB:X05558; NID:g63685; PIDN:CAA29073.1; PID:g63686
C;Genetics:
A;Introns: 355/3; 397/2
C;Superfamily: cytoskeletal keratin
C;Keywords: coiled coil


  Query Match             12.6%;  Score 79;  DB 2;  Length 858;
  Best Local Similarity   25.2%;  Pred. No. 51;
  Matches   36;  Conservative   18;  Mismatches   59;  Indels   30;  Gaps    5;

Qy      3 QKQEEE---------NPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQK------ 47
            :|:|||         : ||| | :|:: :|||       |||: | |    | |
Db    514 EKEEEEAEEEEAAKSDAAEEGGSKKEEIEEKEEGEEAEEEEAEAKGKAEEAGAKVEKVKS 573

Qy     48 ---------PGGSDFL--MKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNL---VT 93
                  |  |      | :||              |:   |:  :| |:|     ||
Db    574 PPAKSPPKSPPKSPVTEQAKAVQKAAAEVGKDQKAEKAAEKAAKEEKAASPEKPATPKVT 633

Qy     94 GDHIP-TPQDLPQRKSSLVTSKL 115
               | ||:  |  : ::   |:
Db    634 SPEKPATPEKPPTPEKAITPEKV 656


RESULT  30
T31433
Na+/Ca2+,K+-exchanging protein - bottle-nosed dolphin
C;Species: Tursiops truncatus (bottle-nosed dolphin)
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T31433
R;Cooper, C.B.; Winkfein, R.J.; Schnetkamp, P.P.M.
submitted to the EMBL Data Library, April 1998
A;Description: Cloning, sequencing, and functional expression of the bottlenose dolph
A;Reference number: Z21032
A;Accession: T31433
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1014 <COO>
A;Cross-references: EMBL:AF059031; NID:g3170602; PID:g3170603; PIDN:AAC18119.1
C;Genetics:
A;Gene: NCKXr


  Query Match             12.6%;  Score 79;  DB 2;  Length 1014;
  Best Local Similarity   29.3%;  Pred. No. 61;
  Matches   22;  Conservative   14;  Mismatches   25;  Indels   14;  Gaps    2;

Qy      4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLM-------- 55
            : |:|   || |||::: :|:|    |  ||  |  ::|   ||     ||:
Db    779 ESEDEEEEEEEGEEEEQEEEEEEQEEEEEEEPLSLEWPETRQKQAIYLFLLPIVFPLWL 838

Qy     56 -----KRLQKGQKYF 65
                :||:  :|:|
Db    839 TVPDVRRLE-ARKFF 852


RESULT  31
S72269
ryanodine receptor isoform 2, cardiac muscle - human
N;Alternate names: calcium release channel protein isoform 2
C;Species: Homo sapiens (man)
C;Date: 14-Apr-1998 #sequence_revision 24-Apr-1998 #text_change 20-Jun-2000
C;Accession: S72269
R;Tunwell, R.E.A.; Wickenden, C.; Bertrand, B.M.A.; Shevchenko, V.I.; Walsh, M.B.; Al
Biochem. J. 318, 477-487, 1996
A;Title: The human cardiac muscle ryanodine receptor-calcium release channel: identif
A;Reference number: S72269; MUID:96404895
A;Accession: S72269
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-4967 <TUN>
A;Cross-references: EMBL:X98330; NID:g1526977; PIDN:CAA66975.1; PID:g1526978
A;Experimental source: cardiac muscle; sarcoplasmic reticulum
C;Genetics:
A;Gene: RyR-2
C;Complex: homotetramer
C;Function:
A;Description: mediates rapid release of Ca2+ ions from intracellular stores during c
```

```
C;Superfamily: ryanodine receptor; transcription initiation factor sigma region 1 homolo
C;Keywords: alternative splicing; calcium channel; cardiac muscle; glycoprotein; heart;
F;710-715/Region: adenine nucleotide binding #status predicted
F;1148-1153/Region: adenine nucleotide binding #status predicted
F;2338-2342/Region: adenine nucleotide binding #status predicted
F;2626-2631/Region: adenine nucleotide binding #status predicted
F;4030-4048/Domain: calcium binding #status predicted <CA1>
F;4499-4519/Domain: transmembrane #status predicted <TM01>
F;4572-4593/Domain: transmembrane #status predicted <TM02>
F;4719-4744/Domain: transmembrane #status predicted <TM03>
F;4767-4789/Domain: transmembrane #status predicted <TM04>
F;4809-4829/Domain: transmembrane #status predicted <TM05>
F;4844-4867/Domain: transmembrane #status predicted <TM06>
F;140,301/Binding site: phosphate (Thr) (covalent) (by cAMP-and calmodulin-dependent kin
F;2808/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status
F;4794/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match           12.6%;  Score 79;  DB 2;  Length 4967;
  Best Local Similarity  23.9%;  Pred. No. 3.5e+02;
  Matches   28;  Conservative   17;  Mismatches   42;  Indels   30;  Gaps    3;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL-----GQKPGGSDFLMKRL 58
           |:||:   |||  | :  : ::|   |:|:| | | |   |     |:          |::
Db    4425 KEEEKEEKEETKSEPEKAEGEDGEKEEKAKEDKGKQKLRQLHTHRYGEPEVPESAFWKKI 4484

Qy      59 ----QKGQKYFDSGDYNM--------------------AKAKMKNKQLPSAGPDKN 90
               ||   ||    |||                    : :: |:||:    :|
Db    4485 IAYQQKLLNYFARNFYNMRMLALFVAFAINFILLFYKVSTSSVVEGKELPTRSSSEN 4541


RESULT  32
S60480
low temperature-induced protein lti29 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 01-Dec-2000
C;Accession: S60480; S52655; S49029
R;Welin, B.V.; Olson, A.; Palva, E.T.
Plant Mol. Biol. 29, 391-395, 1995
A;Title: Structure and organization of two closely related low-temperature-induced dhn/l
A;Reference number: S60480; MUID:96046757
A;Accession: S60480
A;Molecule type: DNA
A;Residues: 1-260 <WEL>
A;Cross-references: EMBL:X90958; NID:g975647; PIDN:CAA62448.1; PID:g975648
R;Welin, B.V.; Olson, A.; Nylander, M.; Palva, E.T.
Plant Mol. Biol. 26, 131-144, 1994
A;Title: Characterization and differential expression of dhn/lea/rab-like genes during c
A;Reference number: S52654; MUID:95035985
A;Accession: S52655
A;Molecule type: mRNA
A;Residues: 'G',57-143,'E',145-260 <WE2>
A;Cross-references: EMBL:X77614; NID:g509261; PIDN:CAA54705.1; PID:g509262
A;Experimental source: ecotype Landsberg erecta
C;Genetics:
A;Gene: lti29; lti45
A;Introns: 110/3


  Query Match           12.6%;  Score 78.5;  DB 2;  Length 260;
  Best Local Similarity  28.4%;  Pred. No. 15;
  Matches   42;  Conservative   19;  Mismatches   42;  Indels   45;  Gaps   11;

Qy       2 SQKQEEENPAEETGEEKQDTQEK---EGILPERAEEA------KLKAKYPSLGQKPGGSD 52
           |   :||    | ||:|:  ::|   ||   :  ||       ::| |:| ||:|||| |
Db     109 SSSSDEEG---EDGEKKKKEKKKKIVEGDHVKTVEEENQGVMDRIKEKFP-LGEKPGGDD 164

Qy      53 FLM------------KRLQKGQKYFDSGDYNMAKAKMKNKQLP--SAGPDKNLVTGD-- 95
             :              |  :: :| |      | | | |   ||  |  |: : |
Db     165 VPVVTTMPAPHSVEDHKPEEEEKKGF------MDKIKEK---LPGHSKKPEDSQVVNTTP 215

Qy      96 --HIPTP-QDLPQRKSSL---VTSKLAG 117
               ||  |:|: |      :  || |


Db    216 LVETATPIADIPEEKKGFMDKIKEKLPG 243


RESULT  33
T34036
hypothetical protein B0041.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34036
R;Fulton, R.; Wohldmann, P.
submitted to the EMBL Data Library, April 1997
A;Description: The sequence of C. elegans cosmid B0041.
A;Reference number: Z21466
A;Accession: T34036
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1359 <FUL>
A;Cross-references: EMBL:AF000196; PIDN:AAC24256.1; GSPDB:GN00019; CESP:B0041.7
A;Experimental source: strain Bristol N2; clone B0041
C;Genetics:
A;Gene: CESP:B0041.7
A;Map position: 1
A;Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3


  Query Match           12.6%;  Score 78.5;  DB 2;  Length 1359;
  Best Local Similarity  26.9%;  Pred. No. 92;
  Matches   32;  Conservative   24;  Mismatches   48;  Indels   15;  Gaps    5;

Qy       2 SQKQEEENPAEETGEE---KQDTQEKEGILPERAE--EAKLKAKYPSLGQKPGGSDFLMK 56
           ::||     :||: ||   |:  : ||  : :|||  |   : : ||   | |     :|
Db     147 TKKQTSSESSEESEEERKVKKSKKNKEKSVKKRAETSEESDEDEKPSKKSKKG-----LK 201

Qy      57 RLQKGQKYFDSGD-YNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSK 114
           :  | :   :| |   : |:| |:|::      |   : |  |  :   :|| |  :|:
Db     202 KKAKSESESESEDEKEVKKSKKKSKKVVK----KESESEDEAPEKKKTEKRKRSKTSSE 256


RESULT  34
S48647
peptidylprolyl isomerase (EC 5.2.1.8) FPR3 - yeast (Saccharomyces cerevisiae)
N;Alternate names: FK506/rapamycin-binding protein FPR3; immunophilin FKBP-70; prolin
C;Species: Saccharomyces cerevisiae
C;Date: 15-Jul-1995 #sequence_revision 09-Mar-1996 #text_change 16-Jul-1999
C;Accession: S48647; S61764; A53544; S47905; S48823; S47927
R;Manning-Krieg, U.C.; Henriquez, R.; Cammas, F.; Graff, P.; Gaveriaux, S.; Movva, N.
FEBS Lett. 352, 98-103, 1994
A;Title: Purification of FKBP-70, a novel immunophilin from Saccharomyces cerevisiae,
A;Reference number: S48647; MUID:95010686
A;Accession: S48647
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-411 <MAN>
A;Cross-references: GB:S73876; NID:g693860; PIDN:AAB31995.1; PID:g693861
A;Accession: S61764
A;Molecule type: protein
A;Residues: 57-70;164-183;303-320;362-379;389-405 <MA2>
R;Benton, B.M.; Zang, J.H.; Thorner, J.
J. Cell Biol. 127, 623-639, 1994
A;Title: A novel FK506- and rapamycin-binding protein (FPR3 gene product) in the yeas
A;Reference number: A53544; MUID:95050937
A;Accession: A53544
A;Molecule type: DNA
A;Residues: 1-121,'P',123-240,'EE',241-334,'F',336-411 <BEN>
A;Cross-references: GB:L34569; NID:g1439536; PIDN:AAB04165.1; PID:g530998
R;Shan, X.; Xue, Z.; Melese, T.
J. Cell Biol. 126, 853-862, 1994
A;Title: Yeast NPI46 encodes a novel prolyl cis-trans isomerase that is located in th
A;Reference number: S47905; MUID:94327703
A;Accession: S47905
A;Molecule type: DNA
A;Residues: 1-240,241-411 <SHA>
A;Cross-references: EMBL:X79379; NID:g531744; PIDN:CAA55924.1; PID:g531745
```

```
R;Brown, D.; Bowman, S.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48816
A;Accession: S48823
A;Molecule type: DNA
A;Residues: 1-240,241-411 <BRO>
A;Cross-references: EMBL:Z46373; NID:g587529; PIDN:CAA86504.1; PID:g587537; MIPS:YML074c
C;Genetics:
A;Gene: SGD:NPI46; FPR3
A;Cross-references: SGD:S0004539; MIPS:YML074c
A;Map position: 13L
C;Function:
A;Description: cis-trans-isomerase
C;Superfamily: yeast peptidylprolyl isomerase FPR3; BKBP-type peptidylprolyl isomerase h
C;Keywords: cis-trans-isomerase; cyclosporin A binding; nucleolus; nucleus
F;324-370/Domain: BKBP-type peptidylprolyl isomerase homology <PPI>


  Query Match             12.5%;  Score 78;  DB 2;  Length 411;
  Best Local Similarity   24.8%;  Pred. No. 27;
  Matches   32;  Conservative   21;  Mismatches   60;  Indels   16;  Gaps    3;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           : :: ||:|  ||  ||::: ::|| : ||  :  | | :     ::         |: :|
Db     219 VQEEDEEDNDGEEEQEEEEEEEQKEEVKPEPKKSKKEKKRKHEEKEEE-------KKAKK 271

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHI----PTPQDLPQRKSSL-----V 111
            :|     |      | |:|:       |: |    |     |  | || |
Db     272 VKKVEFKKDLEEGPTKPKSKKEQDKHKPKSKVLEGGIVIEDRTIGDGPQAKRGARVGMRY 331

Qy     112 TSKLAGGQV 120
             ||  |:|
Db     332 IGKLKNGKV 340


RESULT  35
T41384
hypothetical protein SPCC550.09 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000
C;Accession: T41384
R;Murphy, L.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, May 1998
A;Reference number: Z21990
A;Accession: T41384
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-535 <MUR>
A;Cross-references: EMBL:AL023592; PIDN:CAA19113.1; GSPDB:GN00068; SPDB:SPCC550.09
A;Experimental source: strain 972h-; cosmid c550
C;Genetics:
A;Gene: SPDB:SPCC550.09
A;Map position: 3
C;Superfamily: Schizosaccharomyces pombe hypothetical protein SPCC550.09


  Query Match             12.5%;  Score 78;  DB 2;  Length 535;
  Best Local Similarity   23.8%;  Pred. No. 37;
  Matches   20;  Conservative   19;  Mismatches   45;  Indels    0;  Gaps    0;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           : ::|  ::|: |   |    :  |   |   || :  |:     ::   |  : :  :|
Db     150 VKEQQASQSPSVEGENEPAGVKSNEATKPSVTEEKEKNAEQNKRRERLSISHIVTELAKK 209

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPS 84
            :   :||  :  ||:||:|| |
Db     210 DEVSSESGSNEIVNAKLKNEQLDS 233


RESULT  36
T37740
coiled-coil protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T37740
R;Rieger, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 1999
A;Reference number: Z21743
A;Accession: T37740
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-649 <RIE>
A;Cross-references: EMBL:AL035248; PIDN:CAA22848.1; GSPDB:GN00066; SPDB:SPAC167.03c
A;Experimental source: strain 972h-; cosmid c167
C;Genetics:
A;Gene: SPDB:SPAC167.03c
A;Map position: 1


  Query Match             12.5%;  Score 78;  DB 2;  Length 649;
  Best Local Similarity   26.4%;  Pred. No. 45;
  Matches   37;  Conservative   14;  Mismatches   57;  Indels   32;  Gaps    5;

Qy       7 EENPAEETGE--------------EKQDTQEKEGILPERAEEAKLKAKYPSLGQK---- 47
           || | :|  :              |    ||:| |   | :| :||:|:  | ||
Db      32 EEKPQKELSDASVKSSYVDQEQQAYENWKKQEQEEI--NRKKEEELKSKFEKLRQKNERR 89

Qy      48 ---PGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDL- 103
               | :  | : | :     |: |      ||:|  |   |   |    :  |    |
Db      90 RRTQGKT--LAETLAEEDDQIDANDTRAWILKMRNSSLNKNGNGLNFEAKNDAPRRSTLH 147

Qy     104 -----PQRKSSLVTSKLAGG 118
                |   |||   |:| |
Db     148 SQNAGPNVNSSLEGMKIAHG 167


RESULT  37
T48930
hypothetical protein F14L2.50 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Sep-2000
C;Accession: T48930
R;Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z25008
A;Accession: T48930
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-742 <JOR>
A;Cross-references: EMBL:AL353818; GSPDB:GN00061; ATSP:F14L2.50
A;Experimental source: cultivar Columbia; BAC clone F14L2
C;Genetics:
A;Gene: ATSP:F14L2.50
A;Map position: 3
A;Introns: 115/1; 211/3; 336/3; 348/3; 389/3; 425/3; 444/3; 465/3; 546/2; 573/3; 607/
C;Superfamily: Arabidopsis thaliana hypothetical protein T21L8.10


  Query Match             12.5%;  Score 78;  DB 2;  Length 742;
  Best Local Similarity   22.6%;  Pred. No. 53;
  Matches   24;  Conservative   22;  Mismatches   44;  Indels   16;  Gaps    3;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAE------EAKLKAKYPSLGQKPGGSDFLMKR 57
           :::||   |: | |:|   |:|| |  :||      |  ::      |  ||  :   :
Db     319 EEDEERKQEDEGVERQPEAEEEGGLERKAENDNESFEDSIREPNTQYGSYPGDDENTQR- 377

Qy      58 LQKGQKYFDSGDYNMAKAKMKNKQLP-SAGPDKNLVTGDHIPTPQD 102
                | ||  : ::       | |: |: |::: : :   :|
Db     378 --------DVGDELVEESSKDKSPTPRSSTPNFNILSEESLDVQKD 415


RESULT  38
H83736
subtilisin-type alkaline proteinase (EC 3.4.21.-) apr precursor [similarity] - Bacill
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
```

```
C;Accession: H83736
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: H83736
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-792 <STO>
A;Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BAB04415.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: apr
C;Keywords: hydrolase; serine proteinase
F;1-28/Domain: signal sequence #status predicted <SIG>


  Query Match             12.5%;  Score 78;  DB 2;  Length 792;
  Best Local Similarity   26.5%;  Pred. No. 56;
  Matches   31;  Conservative   22;  Mismatches   46;  Indels   18;  Gaps    5;

Qy       3 QKQEEENPAEETGEEKQDTQEKEG------ILPERAEEAKLKA----KYPSLGQKPGGSD 52
           ::: ||:|||   |:  | :| ||        |:  |  :  |    : || |::  ||:
Db     383 EEENEESPAENPQEQPSDGKENEGSEDQGSTPPDEEENEESPAEDPQEQPSDGKENKGSE 442

Qy      53 FLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSS 109
             :|    |  :   : |::  :| || | : |  : |   || |  : : |
Db     443 ------NQGSTPPDEEENEESPAEVPQEQ-PSDGKE-NEGSEDQGSTPPDEEENEES 491


RESULT  39
I50479
neurofilament medium protein - goldfish
C;Species: Carassius auratus (goldfish)
C;Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 13-Aug-1999
C;Accession: I50479
R;Glasgow, E.; Hall, C.M.; Schechter, N.
J. Neurochem. 63, 52-61, 1994
A;Title: Organization, sequence, and expression of a gene encoding goldfish neurofilamen
A;Reference number: I50479; MUID:94267484
A;Accession: I50479
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-798 <GLA>
A;Cross-references: GB:L09742; NID:g213019; PIDN:AAA72341.1; PID:g213020
C;Genetics:
A;Introns: 343/3; 385/2
C;Superfamily: cytoskeletal keratin


  Query Match             12.4%;  Score 77.5;  DB 2;  Length 798;
  Best Local Similarity   25.4%;  Pred. No. 63;
  Matches   30;  Conservative   17;  Mismatches   54;  Indels   17;  Gaps    4;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           | :||   |  |||::   : |   | ||| |     |      |          :| ::
Db     528 KGDEEEEGEGEGEEEKQEDDAE----EEAEETKA----PEAKASP--------ETEKAEE 571

Qy      64 YFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAGGQVE 121
             ||    |: :   |    || ||    |:  |  :: |::  :  | :   |: |
Db     572 KQTSGGEEEAEEEGDEKDEADAGSDKG-SPGEKEPEAKEQPKKVEAEATKEDKKGKAE 628


RESULT  40
T23648
hypothetical protein M01E5.5a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T23648
R;Dobson, R.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19777
A;Accession: T23648
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-806 <WIL>
A;Cross-references: EMBL:Z93385; PIDN:CAB07640.1; GSPDB:GN00019; CESP:M01E5.5a
A;Experimental source: clone M01E5
C;Genetics:
A;Gene: CESP:M01E5.5a
A;Map position: 1
A;Introns: 133/3; 205/3; 699/3; 780/3
C;Superfamily: eukaryotic type I DNA topoisomerase


  Query Match             12.4%;  Score 77.5;  DB 2;  Length 806;
  Best Local Similarity   29.3%;  Pred. No. 63;
  Matches   36;  Conservative   15;  Mismatches   31;  Indels   41;  Gaps    6;

Qy       6 EEENPAEET-GEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKY 64
           |||: |  |  ||:|  :|||    |: |: : | |               |||:|  :
Db     141 EEEDEAPTTDDEEEQKRKEKERRKKEKREKKERKEK--------------KRLEKENRK 185

Qy      65 F------DSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAGG 118
             || | :  ||| | ::  | | :|:                | |  : | |||
Db     186 IKEEDDEDSDDEDDEKAKKKKRK--SKGAEKS--------------KPS-TSKKDAGG 226

Qy     119 QVE 121
           : |
Db     227 KKE 229


Search completed: September  7, 2002, 10:44:27
Job time: 163 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:50:54 ; Search time 43.01 Seconds
(without alignments)
486.687 Million cell updates/sec

Title: US-09-824-178-4
Perfect score: 625
Sequence: 1 MSQKQEEENPAEETGEEKQD.....DLPQRKSSLVTSKLAGGQVE 121

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	625	100.0	121	4	O43768	O43768 homo sapien
2	622	99.5	121	6	O97976	O97976 sus scrofa
3	614	98.2	121	11	Q9Z0G1	Q9z0g1 mus musculu
4	605	96.8	117	4	Q9NRZ0	Q9nrz0 homo sapien
5	590	94.4	117	11	Q9CQZ9	Q9cqz9 mus musculu
6	400	64.0	110	6	P79322	P79322 sus scrofa
7	226	36.2	119	5	Q9VUB8	Q9vub8 drosophila
8	128.5	20.6	174	5	Q9XU56	Q9xu56 caenorhabdi
9	122	19.5	139	3	P79058	P79058 schizosacch
10	102	16.3	605	4	Q9NZ53	Q9nz53 homo sapien
11	100.5	16.1	417	2	Q9LAY3	Q9lay3 streptococc
12	95.5	15.3	406	2	Q9LAZ0	Q9laz0 streptococc
13	95.5	15.3	842	16	Q98Q49	Q98q49 mycoplasma
14	94.5	15.1	394	2	Q9LAY6	Q9lay6 streptococc
15	94.5	15.1	395	2	Q9LAZ1	Q9laz1 streptococc
16	91.5	14.6	2664	5	Q26033	Q26033 plasmodium
17	91.5	14.6	5120	13	Q9PU36	Q9pu36 gallus gall
18	90.5	14.5	990	4	Q9NYF9	Q9nyf9 homo sapien
19	89	14.2	1165	4	Q9UPN4	Q9upn4 homo sapien
20	88	14.1	1270	4	Q96JH2	Q96jh2 homo sapien
21	87.5	14.0	1751	5	Q9VR59	Q9vr59 drosophila
22	86	13.8	399	10	Q94IK2	Q94ik2 solanum tub
23	85	13.6	775	10	Q9SKL9	Q9skl9 arabidopsis
24	84.5	13.5	967	11	Q920Q9	Q920q9 rattus norv
25	84	13.4	969	5	Q9VLL1	Q9vll1 drosophila
26	84	13.4	969	5	Q95R51	Q95r51 drosophila
27	84	13.4	2242	4	Q9P2P9	Q9p2p9 homo sapien
28	83.5	13.4	281	5	Q9VYC3	Q9vyc3 drosophila
29	83	13.3	131	3	Q9P305	Q9p305 saccharomyc
30	83	13.3	206	11	Q9CUW6	Q9cuw6 mus musculu
31	83	13.3	487	13	O73793	O73793 serinus can
32	83	13.3	565	5	Q960P3	Q960p3 drosophila
33	83	13.3	619	2	Q54972	Q54972 streptococc
34	83	13.3	735	5	Q9U0D5	Q9u0d5 tetrahymena
35	83	13.3	763	10	Q9SUA1	Q9sua1 arabidopsis
36	83	13.3	792	10	Q9SKM3	Q9skm3 arabidopsis
37	83	13.3	2274	5	Q9VYU0	Q9vyu0 drosophila
38	82.5	13.2	795	10	Q9XGD3	Q9xgd3 arabidopsis
39	82	13.1	335	5	O76719	O76719 caenorhabdi
40	82	13.1	2768	5	Q9VC00	Q9vc00 drosophila
41	81.5	13.0	646	10	O64877	O64877 arabidopsis
42	81.5	13.0	789	10	Q9SI46	Q9si46 arabidopsis
43	81.5	13.0	806	5	Q27529	Q27529 caenorhabdi
44	81.5	13.0	893	4	Q9Y5T0	Q9y5t0 homo sapien
45	81.5	13.0	2817	4	Q96P79	Q96p79 homo sapien

ALIGNMENTS

RESULT 1
O43768
ID O43768 PRELIMINARY; PRT; 121 AA.
AC O43768;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ALPHA-ENDOSULFINE.
GN ENSA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND PHOSPHORYLATION.
RC TISSUE=BRAIN;
RX MEDLINE=98318667; PubMed=9653196;
RA Heron L., Virsolvy A., Peyrollier K., Gribble F.M., Le Cam A.,
RA Ashcroft F.M., Bataille D.;
RT "Human alpha-endosulfine, a possible regulator of sulfonylurea-
RT sensitive K(ATP) channel: molecular cloning, expression and biological
RT properties.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8387-8391(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99408499; PubMed=10480622;
RA Heron L., Virsolvy A., Apiou F., Le Cam A., Bataille D.;
RT "Isolation, characterization, and chromosomal localization of the
RT human ENSA gene that encodes alpha-endosulfine, a regulator of beta-
RT cell K(ATP) channels.";
RL Diabetes 48:1873-1876(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Zhang Q., Fu G., Wu J., Zhou J., Ye M., Shen Y., Kan L., He K., Gu B.,
RA Chen S., Mao M., Chen Z.;
RT "Human alpha endosulfine gene.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN [4]

```
RP   SEQUENCE FROM N.A.
RA   Scott V.E.S., Roch J.-M., Davis-Taber R.A., Molinari E.J.,
RA   Whiteaker K.L., Gopalakrishnan M., Idler K., Sullivan J.P.;
RT   "Cloning and molecular characterization of two isoforms of human
RT   endosulfine.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG;
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ENDOGENOUS LIGAND FOR SULFONYLUREA RECEPTOR. BY
CC       INHIBITING SULFONYLUREA FROM BINDING TO THE RECEPTOR, IT REDUCES
CC       K(ATP) CHANNEL CURRENTS AND THEREBY STIMULATES INSULIN SECRETION
CC       (PROBABLE).
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGH LEVELS IN MUSCLE
CC       AND BRAIN AND LOWER LEVELS IN THE PANCREAS.
CC   -!- PTM: PHOSPHORYLATED BY PROTEIN KINASE A.
CC   -!- SIMILARITY: STRONG, TO ARPP-19.
DR   EMBL; X99906; CAA68180.1; -.
DR   EMBL; AJ010966; CAB65125.1; -.
DR   EMBL; AF067170; AAD32454.1; -.
DR   EMBL; AF157509; AAF80340.1; -.
DR   EMBL; BC004461; AAH04461.1; -.
DR   EMBL; BC000436; AAH00436.1; -.
DR   MIM; 603061; -.
KW   Phosphorylation.
SQ   SEQUENCE   121 AA;  13389 MW;  7C76315AA17E7542 CRC64;
```

```
  Query Match           100.0%;  Score 625;  DB 4;  Length 121;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-50;
  Matches  121;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAGGQV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAGGQV 120

Qy     121 E 121
           |
Db     121 E 121
```

```
RESULT   2
O97976
ID   O97976      PRELIMINARY;      PRT;   121 AA.
AC   O97976; Q95105;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA-ENDOSULFINE.
OS   Sus scrofa (Pig), and
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823, 9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Pig, and Bovine; TISSUE=BRAIN;
RA   Bataille D.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 25-101 FROM N.A.
RC   SPECIES=Bovine; TISSUE=CAUDATE;
RX   MEDLINE=96268722; PubMed=8687439;
RA   Peyrollier K., Heron L., Virsolvy-Vergine A., Le Cam A., Bataille D.;
RT   "Alpha endosulfine is a novel molecule, structually related to a
RT   family of phosphoproteins.";
RL   Biochem. Biophys. Res. Commun. 223:583-586(1996).
```

```
RN   [3]
RP   SEQUENCE OF 25-36; 39-57; 59-74 AND 81-106, AND CHARACTERIZATION.
RC   SPECIES=Pig; TISSUE=BRAIN;
RX   MEDLINE=96218590; PubMed=8635664;
RA   Virsolvy-Vergine A., Salazar G., Sillard R., Denoroy L., Mutt V.,
RA   Bataille D.;
RT   "Endosulfine, endogenous ligand for the sulfonylurea receptor:
RT   isolation and partial structural determination of the alpha form.";
RL   Diabetologia 39:135-141(1996).
CC   -!- FUNCTION: ENDOGENOUS LIGAND FOR SULFONYLUREA RECEPTOR. BY
CC       INHIBITING SULFONYLUREA FROM BINDING TO THE RECEPTOR, IT REDUCES
CC       K(ATP) CHANNEL CURRENTS AND THEREBY STIMULATES INSULIN SECRETION
CC       (PROBABLE).
CC   -!- PTM: PHOSPHORYLATED BY PROTEIN KINASE A (BY SIMILARITY).
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- MASS SPECTROMETRY: MW=13196; METHOD=ELECTROSPRAY.
CC   -!- SIMILARITY: STRONG, TO ARPP-19.
DR   EMBL; AJ005987; CAA06801.1; -.
DR   EMBL; AJ005986; CAA06800.1; -.
DR   EMBL; X95958; CAA65196.1; -.
KW   Phosphorylation.
FT   VARIANT      31     31       K -> R.
SQ   SEQUENCE   121 AA;  13361 MW;  A84629D3B9EFFC5A CRC64;
```

```
  Query Match           99.5%;  Score 622;  DB 6;  Length 121;
  Best Local Similarity  99.2%;  Pred. No. 2.5e-50;
  Matches  120;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||
Db       1 MSQKQEEENPAEETGEEKQDTQEKEGILPEKAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAGGQV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAGGQV 120

Qy     121 E 121
           |
Db     121 E 121
```

```
RESULT   3
Q9Z0G1
ID   Q9Z0G1      PRELIMINARY;      PRT;   121 AA.
AC   Q9Z0G1;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ALPHA-ENDOSULFINE.
GN   ENSA.
OS   Mus musculus (Mouse), and
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090, 10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Mouse, and Rat; TISSUE=BRAIN;
RA   Bataille D.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Mouse; STRAIN=C57BL/6J; TISSUE=TESTIS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
```

```
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- FUNCTION: ENDOGENOUS LIGAND FOR SULFONYLUREA RECEPTOR. BY
CC       INHIBITING SULFONYLUREA FROM BINDING TO THE RECEPTOR, IT REDUCES
CC       K(ATP) CHANNEL CURRENTS AND THEREBY STIMULATES INSULIN SECRETION
CC       (BY SIMILARITY).
CC   -!- PTM: PHOSPHORYLATED BY PROTEIN KINASE A (BY SIMILARITY).
CC   -!- SIMILARITY: STRONG, TO ARPP-19.
DR   EMBL; AJ005985; CAA06799.1; -.
DR   EMBL; AJ005984; CAA06798.1; -.
DR   EMBL; AK006149; BAB24433.1; -.
DR   MGD; MGI:1891189; Ensa.
KW   Phosphorylation.
SQ   SEQUENCE   121 AA;  13335 MW;  A84632C9085FFC5A CRC64;


  Query Match             98.2%;  Score 614;  DB 11;  Length 121;
  Best Local Similarity   98.3%;  Pred. No. 1.4e-49;
  Matches  119;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           |||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db       1 MSQKQEEENPAEETGEEKQDTQEKEGILPEKAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAGGQV 120
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      61 GQKYFDSGDYNMAKAKMKNKQLPSAGADKNLVTGDHIPTPQDLPQRKSSLVTSKLAGGQV 120

Qy     121 E 121
           |
Db     121 E 121


RESULT   4
Q9NRZ0
ID   Q9NRZ0      PRELIMINARY;      PRT;   117 AA.
AC   Q9NRZ0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   BETA ENDOSULFINE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Scott V.E.S., Roch J.-M., Davis-Taber R.A., Molinari E.J.,
RA   Whiteaker K.L., Gopalakrishnan M., Idler K., Sullivan J.P.;
RT   "Cloning and molecular characterization of two isoforms of human
RT   endosulfine.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF157510; AAF80341.1; -.
SQ   SEQUENCE   117 AA;  12976 MW;  D17E754219767472 CRC64;


  Query Match             96.8%;  Score 605;  DB 4;  Length 117;
  Best Local Similarity   100.0%;  Pred. No. 9e-49;
  Matches  117;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117


RESULT   5
Q9CQZ9
ID   Q9CQZ9      PRELIMINARY;      PRT;   117 AA.
AC   Q9CQZ9;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   2610007F17RIK PROTEIN.
GN   ENSA OR 2610007F17RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK019170; BAB31584.1; -.
DR   EMBL; AK011334; BAB27552.1; -.
DR   MGD; MGI:1891189; Ensa.
SQ   SEQUENCE   117 AA;  12922 MW;  3AD5D4F0B07898C8 CRC64;


  Query Match             94.4%;  Score 590;  DB 11;  Length 117;
  Best Local Similarity   97.4%;  Pred. No. 2.2e-47;
  Matches  114;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           |||||||||||||||||||||:||||||:||||||||||||||||||||||||||||||
Db       1 MSQKQEEENPAEETGEEKQDTQEEEGILPEKAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      61 GQKYFDSGDYNMAKAKMKNKQLPSAGADKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117


RESULT   6
P79322
ID   P79322      PRELIMINARY;      PRT;   110 AA.
AC   P79322;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CAMP-REGULATED PHOSPHOPROTEIN 19K (FRAGMENT).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
```

```
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SMALL INTESTINE;
RA   Winteroe A.K., Fredholm M.;
RT   "Evaluation and characterization of a porcine small intestine cDNA
RT   library.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z84165; CAB06328.1; -.
FT   NON_TER      1      1
SQ   SEQUENCE   110 AA;  12189 MW;  DF42271D5CC114D3 CRC64;


  Query Match             64.0%;  Score 400;  DB 6;  Length 110;
  Best Local Similarity   77.2%;  Pred. No. 8e-30;
  Matches   78;  Conservative    6;  Mismatches   17;  Indels    0;  Gaps    0;

Qy      17 EKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAK 76
           |:|   | :   ||:|||||||:|| |||||||||| ||||||||||||||||||||
Db       5 EEQKEMEDKVTSPEKAEEAKLKARYPHLGQKPGGSDFLRKRLQKGQKYFDSGDYNMAKAK 64

Qy      77 MKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           | |||||:| |||  |||||||  :|||||| ||| |||||
Db      65 MXNKQLPTAAPDKTXVTGDHIPXHRDLPQRKPSLVASKLAG 105


RESULT   7
Q9VUB8
ID   Q9VUB8      PRELIMINARY;      PRT;   119 AA.
AC   Q9VUB8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG6513 PROTEIN.
GN   CG6513.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003536; AAF49770.1; -.
DR   FlyBase; FBgn0036401; CG6513.
SQ   SEQUENCE   119 AA;  12949 MW;  B5B328A1CBF9E69C CRC64;


  Query Match             36.2%;  Score 226;  DB 5;  Length 119;
  Best Local Similarity   43.7%;  Pred. No. 1.1e-13;
  Matches   52;  Conservative   17;  Mismatches   32;  Indels   18;  Gaps    3;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGG-SDFLMKRLQ 59
           ||  :|  |    | :: : |::      |: || |||:||||  : ||| | || ||||
Db       1 MSSAEENSNSPATTPQDTETTEQANLTDLEKIEEEKLKSKYPSGMRVPGGHSAFLQKRLQ 60

Qy      60 KGQKYFDSGDYNMAKAK-------MKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLV 111
           ||||:|||||| ||| |          ||          : ||: |||| : || ||:|::
Db      61 KGQKFFDSGDYQMAKQKGGGVKQVFANK----------VTTGEAIPTPETVPARKTSII 109


RESULT   8
Q9XU56
ID   Q9XU56      PRELIMINARY;      PRT;   174 AA.
AC   Q9XU56;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   K10C3.2 PROTEIN.
GN   K10C3.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mortimore B.J.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z83235; CAB05775.1; -.
SQ   SEQUENCE   174 AA;  18108 MW;  64EB4E9176F4B36E CRC64;


  Query Match             20.6%;  Score 128.5;  DB 5;  Length 174;
  Best Local Similarity   29.0%;  Pred. No. 0.00019;
  Matches   38;  Conservative   16;  Mismatches   24;  Indels   53;  Gaps    5;

Qy      29 PERAEEAKLKAKYPSLGQKPG--GSDFLMKRLQKGQKYFDSGDYNMAKAK----MKNKQL 82
           ||: :| :|  |  : |: |    | || |:||: :|:|||||| | |:|    : :|
Db      22 PEKQQEQELMGKLAATGKLPARPASSFLQKKLQQ-RKFFDSGDYAMDKSKAGTGLGSKPH 80

Qy      83 PSAG-----------------------------------------PDKNLVTGDHIPTP 100
           | ||                                         | ||     || |
Db      81 PLAGGPPPAAPPVVAQRSPAPAATTPSPSASPISQQTNRPSSDRNSDDDNL----QIPRP 136

Qy     101 QDLPQRKSSLV 111
             :||||:|::
Db     137 DTVPQRKASII 147


RESULT   9
```

```
P79058
ID   P79058      PRELIMINARY;      PRT;   139 AA.
AC   P79058;
DT   01-JAN-1999 (TrEMBLrel. 09, Created)
DT   01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   HYPOTHETICAL 15.4 KDA PROTEIN C10F6.16 IN CHROMOSOME I.
GN   SPAC10F6.16.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 20-139 FROM N.A.
RA   Kawamukai M.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO YEAST YNL157W.
DR   EMBL; AL009197; CAA15729.1; -.
DR   EMBL; AB001289; BAA19234.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   139 AA;  15441 MW;  D8D9D2A388ADA532 CRC64;


  Query Match             19.5%;  Score 122;  DB 3;  Length 139;
  Best Local Similarity   34.1%;  Pred. No. 0.00058;
  Matches   30;  Conservative   15;  Mismatches   31;  Indels   12;  Gaps    3;

Qy      20 DTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKN 79
           ::::|  :    || ||   |   |: |   | |:::||:|:|||||||| : ||
Db      20 NSEQKVDVAKLSPEEQKLFRLY---GRLPQRKDLLVQKLQQGRKYFDSGDYALNKA---- 72

Qy      80 KQLPSAGPDKNLV-TGDHIPTPQDLPQR 106
                 |  :   |  ||:|  :| |
Db      73 ----GKASDSGITCIGKEIPSPDTIPHR 96


RESULT  10
Q9NZ53
ID   Q9NZ53      PRELIMINARY;      PRT;   605 AA.
AC   Q9NZ53;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ENDOGLYCAN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=20187625; PubMed=10722749;
RA   Sassetti C., Van Zante A., Rosen S.D.;
RT   "Identification of Endoglycan, a Member of the CD34/Podocalyxin Family
RT   of Sialomucins.";
RL   J. Biol. Chem. 275:9001-9010(2000).
DR   EMBL; AF219137; AAF44629.1; -.
SQ   SEQUENCE   605 AA;  65075 MW;  2F9B8DC51F7FC22A CRC64;


  Query Match             16.3%;  Score 102;  DB 4;  Length 605;
  Best Local Similarity   27.9%;  Pred. No. 0.22;
  Matches   41;  Conservative   21;  Mismatches   47;  Indels   38;  Gaps    7;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERA--EEAKLKAKYPSL---GQKPGGSDFLMKR 57
           :::|||   |:   |||: :|:| :||     |||| : :  ||    | || :
Db     165 EEEEEEEEREKEEVEKQEEEEEEELLPVNGSQEEAKPQVRDFSLTSSSQTPGATK----- 219

Qy      58 LQKGQKYFDSGDYNMAKAKMKNKQ-----LPSAGPDKNLVTGDHIPTPQ---------- 101
              :: ||||   :  ::::      |||  |   :  ||   | |
Db     220 ----SRHEDSGDQASSGVEVESSMGPSLLLPSVTP-TTVTPGDQDSTSQEAEATVLPAAG 274

Qy     102 -----DLPQRKSSLVTSKLAG--GQVE 121
               : ||  |   |:  ||  || |
Db     275 LGVEFEAPQEASEEATAGAAGLSGQHE 301


RESULT  11
Q9LAY3
ID   Q9LAY3      PRELIMINARY;      PRT;   417 AA.
AC   Q9LAY3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PSPA (FRAGMENT).
GN   PSPA.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EF10197;
RX   MEDLINE=20448953; PubMed=10992499;
RA   Hollingshead S.K., Becker R., Briles D.E.;
RT   "Diversity of PspA: mosaic genes and evidence for past recombination
RT   in Streptococcus pneumoniae.";
RL   Infect. Immun. 68:5889-5900(2000).
DR   EMBL; AF071812; AAF27708.1; -.
DR   HSSP; P00192; 1APC.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR001854; Ribosomal_L29.
DR   PRINTS; PR01217; PRICHEXTENSN.
FT   NON_TER     417    417
SQ   SEQUENCE   417 AA;  46960 MW;  876EAD3276506EEC CRC64;


  Query Match             16.1%;  Score 100.5;  DB 2;  Length 417;
  Best Local Similarity   29.4%;  Pred. No. 0.2;
  Matches   37;  Conservative   20;  Mismatches   48;  Indels   21;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           |:: || |::  || :  |||       | ||   | :  || |:    |: :|     :|
Db      50 KRDAEN-AKKALEEAKRAQEKYADYQRRIEEKAAKEQQASLEQQEANKDYQLKL----KK 104

Qy      64 YFD----------SGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDL--PQRKSSLV 111
           | |            :   |: | | ||       : :|    :| ||:|   :||| :|
Db     105 YLDGRNLSNSSVLKKEMEEAEKKDKEKQAEFNKIRREIV----VPNPQELEMARRKSEVV 160

Qy     112 TSKLAG 117
             :| :|
Db     161 KAKESG 166


RESULT  12
Q9LAZ0
ID   Q9LAZ0      PRELIMINARY;      PRT;   406 AA.
AC   Q9LAZ0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PSPA (FRAGMENT).
GN   PSPA.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DBL6A;
```

```
RX   MEDLINE=20448953; PubMed=10992499;
RA   Hollingshead S.K., Becker R., Briles D.E.;
RT   "Diversity of PspA: mosaic genes and evidence for past recombination
RT   in Streptococcus pneumoniae.";
RL   Infect. Immun. 68:5889-5900(2000).
DR   EMBL; AF071805; AAF27701.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR000533; Tropomyosin.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR00194; TROPOMYOSIN.
FT   NON_TER     406    406
SQ   SEQUENCE   406 AA;  45406 MW;  5B1E20CA4D06F052 CRC64;
```

```
  Query Match             15.3%;  Score 95.5;  DB 2;  Length 406;
  Best Local Similarity   28.6%;  Pred. No. 0.57;
  Matches   36;  Conservative   19;  Mismatches   50;  Indels   21;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           |:: || |::  || :  |||      | ||   |  : || |:    |: :|     :|
Db      50 KRDAEN-AKKALEEAKRAQEKYADYQRRIEEKAAKETHASLEQQEANKDYQLKL----KK 104

Qy      64 YFD----------SGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDL--PQRKSSLV 111
           | |            :   |: | | ||       : :|    :| ||:|   :||| :
Db     105 YLDGRNLSNSSVLKKEMEEAEKKDKEKQAEFNKIRREIV----VPNPQELEMARRKSEVA 160

Qy     112 TSKLAG 117
             | :|
Db     161 KPKESG 166
```

```
RESULT  13
Q98Q49
ID   Q98Q49      PRELIMINARY;      PRT;   842 AA.
AC   Q98Q49;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   HEAT SHOCK ATP-DEPENDENT PROTEASE (EC 3.4.21.53).
GN   MYPU_5190.
OS   Mycoplasma pulmonis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2107;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UAB CTIP;
RX   MEDLINE=21267165; PubMed=11353084;
RA   Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA   Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA   Blanchard A.;
RT   "The complete genome sequence of the murine respiratory pathogen
RT   Mycoplasma pulmonis.";
RL   Nucleic Acids Res. 29:2145-2153(2001).
DR   EMBL; AL445564; CAC13692.1; -.
DR   MypuList; MYPU_5190; -.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003959; AAA_subfam.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   InterPro; IPR003111; LON.
DR   InterPro; IPR001984; Lon_endopep.
DR   Pfam; PF00004; AAA; 1.
DR   PRINTS; PR00830; ENDOLAPTASE.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00464; LON; 1.
KW   Hydrolase; Complete proteome.
SQ   SEQUENCE   842 AA;  95888 MW;  99A276B296EDBFC2 CRC64;
```

```
  Query Match             15.3%;  Score 95.5;  DB 16;  Length 842;
  Best Local Similarity   31.5%;  Pred. No. 1.3;
  Matches   35;  Conservative   19;  Mismatches   48;  Indels    9;  Gaps    4;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQ-KPGGSDFLMKRLQKG 61
           : |: ||  :|  ::     |:|| || |     |:||    ||:      ::  |   |
Db     234 EDQDLENAIDEKMKKNLSKQQKEFILRE-----KMKAIRSELGESNDDEEEYKTKIKDKN 288

Qy      62 QK--YFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSL 110
           |   |  |    :|  : | | : |: || | :| ::| |   || ||:|:
Db     289 NKKIYPKSVRKLIANEESKLKNMMSSSPDAN-ITRNYIDTLMKLPWRKTSI 338
```

```
RESULT  14
Q9LAY6
ID   Q9LAY6      PRELIMINARY;      PRT;   394 AA.
AC   Q9LAY6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PSPA (FRAGMENT).
GN   PSPA.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=L81905;
RX   MEDLINE=20448953; PubMed=10992499;
RA   Hollingshead S.K., Becker R., Briles D.E.;
RT   "Diversity of PspA: mosaic genes and evidence for past recombination
RT   in Streptococcus pneumoniae.";
RL   Infect. Immun. 68:5889-5900(2000).
DR   EMBL; AF071809; AAF27705.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR000533; Tropomyosin.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR00194; TROPOMYOSIN.
FT   NON_TER     394    394
SQ   SEQUENCE   394 AA;  42935 MW;  A192F47E12C88FAA CRC64;
```

```
  Query Match             15.1%;  Score 94.5;  DB 2;  Length 394;
  Best Local Similarity   28.6%;  Pred. No. 0.68;
  Matches   36;  Conservative   19;  Mismatches   50;  Indels   21;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           |:: || |::  || :  |||      | ||   |    || |:    |: :|     :|
Db      50 KRDAEN-AKKALEEAKRAQEKYADYQRRIEEKAAKETQASLEQQEANKDYQLKL----KK 104

Qy      64 YFD----------SGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDL--PQRKSSLV 111
           | |            :   |: | | ||       : :|    :| ||:|   :||| :
Db     105 YLDGRNLSNSSVLKKEMEEAEKKDKEKQAEFNKIRREIV----VPNPQELEMARRKSEVA 160

Qy     112 TSKLAG 117
             :| :|
Db     161 KAKESG 166
```

```
RESULT  15
Q9LAZ1
ID   Q9LAZ1      PRELIMINARY;      PRT;   395 AA.
AC   Q9LAZ1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PSPA (FRAGMENT).
GN   PSPA.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BG9739;
RX   MEDLINE=20448953; PubMed=10992499;
```

```
RA   Hollingshead S.K., Becker R., Briles D.E.;
RT   "Diversity of PspA: mosaic genes and evidence for past recombination
RT   in Streptococcus pneumoniae.";
RL   Infect. Immun. 68:5889-5900(2000).
DR   EMBL; AF071804; AAF27700.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR000533; Tropomyosin.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR00194; TROPOMYOSIN.
FT   NON_TER     395    395
SQ   SEQUENCE   395 AA;  42968 MW;  67FDECAC41DB7F95 CRC64;


  Query Match             15.1%;  Score 94.5;  DB 2;  Length 395;
  Best Local Similarity   28.6%;  Pred. No. 0.68;
  Matches   36;  Conservative   19;  Mismatches   50;  Indels   21;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           |:: || |::  || :  |||      | ||   |    || |:    |: :|     :|
Db      50 KRDAEN-AKKALEEAKRAQEKYADYQRRIEEKAAKETQASLEQQEANKDYQLKL----KK 104

Qy      64 YFD----------SGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDL--PQRKSSLV 111
           | |            :   |: | | ||       : :|    :| ||:|   :||| :
Db     105 YLDGRNLSNSSVLKKEMEEAEKKDKEKQAEFNKIRREIV----VPNPQELEMARRKSEVA 160

Qy     112 TSKLAG 117
             :| :|
Db     161 KAKESG 166


RESULT  16
Q26033
ID   Q26033      PRELIMINARY;      PRT;  2664 AA.
AC   Q26033;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   VARIANT-SPECIFIC SURFACE PROTEIN.
GN   VAR-2.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FCR3;
RX   MEDLINE=95330813; PubMed=7606788;
RA   Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
RA   Peterson D.S., Ravetch J.A., Wellems T.E.;
RT   "The large diverse gene family var encodes proteins involved in
RT   cytoadherence and antigenic variation of Plasmodium falciparum-
RT   infected erythrocytes.";
RL   Cell 82:89-100(1995).
DR   EMBL; L40609; AAA75398.1; -.
DR   InterPro; IPR004258; PFEMP.
DR   Pfam; PF03011; PFEMP; 2.
SQ   SEQUENCE   2664 AA;  302410 MW;  6EA2468511703091 CRC64;


  Query Match             14.6%;  Score 91.5;  DB 5;  Length 2664;
  Best Local Similarity   26.8%;  Pred. No. 12;
  Matches   34;  Conservative   16;  Mismatches   30;  Indels   47;  Gaps    5;

Qy       6 EEENPAEETG-----EEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           ||||| |: |     ::: :  :| | | |: :| | : : |:  :: ||:
Db    2086 EEENPVEQPGFCPTPQQEPEPDDKCGKLEEKKDEKKEQPEQPA--EEDGGA--------- 2134

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGP------DKNLVTGDHIPTPQDLPQRK-----SS 109
                              :||  |      ||  |    || ||: |          |
Db    2135 -------------------IVPSGPPGSEPEADKGPVKPAEIPKPQEPPDLSHPAVIPS 2174

Qy     110 LVTSKLA 116
           |||| ||
Db    2175 LVTSTLA 2181


RESULT  17
Q9PU36
ID   Q9PU36      PRELIMINARY;      PRT;  5120 AA.
AC   Q9PU36;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ACZONIN (FRAGMENT).
GN   ACZ.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99439764; PubMed=10508862;
RA   Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA   Kilimann M.W.;
RT   "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT   zones, shares homology regions with rim and bassoon and binds
RT   profilin.";
RL   J. Cell Biol. 147:151-162(1999).
DR   EMBL; Y19187; CAB60725.1; -.
DR   HSSP; P04410; 1A25.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR001565; Synaptotagmin.
DR   Pfam; PF00168; C2; 2.
DR   Pfam; PF00595; PDZ; 1.
DR   PRINTS; PR00360; C2DOMAIN.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR00399; SYNAPTOTAGMN.
DR   SMART; SM00239; C2; 2.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS00499; C2_DOMAIN_1; UNKNOWN_1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 2.
DR   PROSITE; PS50106; PDZ; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   5120 AA;  560751 MW;  A658D9891B65B412 CRC64;


  Query Match             14.6%;  Score 91.5;  DB 13;  Length 5120;
  Best Local Similarity   21.8%;  Pred. No. 25;
  Matches   32;  Conservative   26;  Mismatches   48;  Indels   41;  Gaps    3;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGIL---------PERAEEAKLKAKYPSLGQKPGGSDF 53
           || | | | ::  |:||:  :|| :          ||:||| |       | :    ||
Db    1194 QKVEVEAPCDKLHEKKQEDVKKEDLTTGIPQMVSKPEKAEEEKTPVPVSRLPR----SDH 1249

Qy      54 LMKRLQKGQKYFDSGDYNMAKAKMK---------------------------NKQLPSA 85
           :    :| :|  |  | : :: :                             | ||
Db    1250 VEAVREKIEKEDDKSDTSSSQQQKSPQGLSDTGYSSDGISSSLGEIPSHIPSDEKDLPRE 1309

Qy      86 GPDKNLVTGDHIPTPQDLPQRKSSLVT 112
             |: :: :  |:| || : :|::::
Db    1310 PSQKDTISQESPPSPSDLAKLESTVLS 1336


RESULT  18
Q9NYF9
ID   Q9NYF9      PRELIMINARY;      PRT;   990 AA.
AC   Q9NYF9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   GOLGI MATRIX PROTEIN GM130.
GN   GOLGA2.
OS   Homo sapiens (Human).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   Weide T.B., Bayer M., Barnekow A.;
RT   "Human golgi matrix protein 130, huGM130, a protein highly homologous
RT   to GM130 from rat.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF248953; AAF65550.1; -.
KW   Matrix protein.
SQ   SEQUENCE   990 AA;  111658 MW;  2380FBFD11B8DAF5 CRC64;


  Query Match             14.5%;  Score 90.5;  DB 4;  Length 990;
  Best Local Similarity   26.8%;  Pred. No. 4.6;
  Matches   40;  Conservative   23;  Mismatches   51;  Indels   35;  Gaps    6;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPER------------AEEAKLKAKYP----SLGQK 47
             ||: :   ::  |||::  :|:| | |:           :|:|:|:      :  ||
Db     176 KQQNQEITDQLEEEKKECHQKQGALREQLQVHIQTIGILVSEKAELQTALAHTQHAARQK 235

Qy      48 PGGSDFLMKRLQKGQKYFDSGDYNMA---------KAKMKNKQLPSAGPDKNLVTGDHIP 98
           | |: |  |||  ::    |:   |          ||   ||:|        |    :
Db     236 EGESEDLASRLQYSRRRV--GELERALSAVSTQQKKADRYNKELTKERDALRLELYKNTQ 293

Qy      99 TPQDLPQRKSS-------LVTSKLAGGQV 120
           : :|| | ||        ||| | || |:
Db     294 SNEDLKQEKSELEEKLRVLVTEK-AGMQL 321


RESULT  19
Q9UPN4
ID   Q9UPN4      PRELIMINARY;      PRT;  1165 AA.
AC   Q9UPN4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA1118 PROTEIN (FRAGMENT).
GN   KIAA1118.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99397452; PubMed=10470851;
RA   Kikuno R., Nagase T., Ishikawa K., Hirosawa M., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIV.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:197-205(1999).
DR   EMBL; AB029041; BAA83070.1; -.
DR   InterPro; IPR000048; IQ.
DR   InterPro; IPR001978; Troponin.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF00992; Troponin; 1.
DR   SMART; SM00015; IQ; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   1165 AA;  130214 MW;  D5517BFB825FDF1A CRC64;


  Query Match             14.2%;  Score 89;  DB 4;  Length 1165;
  Best Local Similarity   23.2%;  Pred. No. 7.6;
  Matches   33;  Conservative   19;  Mismatches   48;  Indels   42;  Gaps    5;

Qy       5 QEEENPAEETGEEK---------QDTQEKEGILPERAEEAK--------------LKAKY 41
           |::|    :  |||         |: |:|  :  ::|  |:              ::
Db     400 QQKEAARRKAREEKARQARRAAIQELQQKRALRAQKASTAERGPPENPRETRVPGMRQPA 459

Qy      42 PSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGP-DKNLVTGDHIPTP 100
             |   |||:                  :   ||      ||:||| |: | | |  | |
Db     460 QELSPTPGGT----------------AHQALKANNAGGGLPAAGPGDRCLPTSDSSPEP 502

Qy     101 QDLPQ-RKSSLVTSKLAGGQVE 121
           |  |: |   ::    ||  :|
Db     503 QQPPEDRTQDVLAQDAAGDNLE 524


RESULT  20
Q96JH2
ID   Q96JH2      PRELIMINARY;      PRT;  1270 AA.
AC   Q96JH2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA1855 PROTEIN (FRAGMENT).
GN   KIAA1855.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=21245130; PubMed=11347906;
RA   Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XX.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large Proteins in vitro.";
RL   DNA Res. 8:85-95(2001).
DR   EMBL; AB058758; BAB47484.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   1270 AA;  137310 MW;  61D4AB45414E3FEC CRC64;


  Query Match             14.1%;  Score 88;  DB 4;  Length 1270;
  Best Local Similarity   30.0%;  Pred. No. 10;
  Matches   36;  Conservative   18;  Mismatches   52;  Indels   14;  Gaps    4;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPG----GSDFLMKRLQ 59
           ::|||   ||  ||::: :|:|    |  | |   |   || : |    ||:    | |
Db     691 EEEEEEDEEEEEEEEEEEEEEEEEEEEEEEEAAAAVALGEVLGPRSGSSSEGSERSTDRSQ 750

Qy      60 KGQK---YFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLA 116
           :|       |    :  :| | :  |       | ||    : :| |:   | | ||::||
Db     751 EGAPSTLLADDQKESRGRASMADGDLEPEEGSKTLV----LVSPGDM---KKSPVTAELA 803


RESULT  21
Q9VR59
ID   Q9VR59      PRELIMINARY;      PRT;  1751 AA.
AC   Q9VR59;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG14616 PROTEIN.
GN   CG14616.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
```

```
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003574; AAF50947.1; -.
DR   FlyBase; FBgn0031192; CG14616.
DR   InterPro; IPR000560; His_acid_phosphtse.
DR   PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; UNKNOWN_1.
SQ   SEQUENCE   1751 AA;  194376 MW;  4556EA0BE7B2C8E4 CRC64;


  Query Match             14.0%;  Score 87.5;  DB 5;  Length 1751;
  Best Local Similarity   26.3%;  Pred. No. 17;
  Matches   35;  Conservative   18;  Mismatches   57;  Indels   23;  Gaps    4;

Qy       3 QKQEEENPA-------EETGEEKQDTQEKEGILPERAEEAKLKAKYPSLG---QKPGGSD 52
           |  :| |||         :|||::   :| |   :|: :   :   |: |    :
Db     976 QSSDESNPARIEEENDSNSGEEREG--KKRGTSGQRSTDRSAERISPAFGFNRLELRSKQ 1033

Qy      53 FLMKRLQKGQKYFDSGDYNMAKAK-----------MKNKQLPSAGPDKNLVTGDHIPTPQ 101
           |  | :  |  :  ||   |  ||            :|:||    ||   || |  |  :
Db   1034 FKSKPIPIGAHHTVSGHEAMDLAKRLNEELASHQQQQNQQLRPISPDIRAVTPDCEPRSR 1093

Qy     102 DLPQRKSSLVTSK 114
           .  || :| | :|
Db   1094 SFEQRPTSGVCAK 1106


RESULT  22
Q94IK2
ID   Q94IK2      PRELIMINARY;      PRT;   399 AA.
AC   Q94IK2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   STOREKEEPER PROTEIN.
GN   STK.
OS   Solanum tuberosum (Potato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4113;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. DESIREE;
RA   Zourelidou M., de Torres Zabala M., Smith C., Bevan M.W.;
RT   "STORE KEEPER (STK) regulates the expression of the potato storage
RT   protein patatin and is the first member of a plant-specific family of
RT   DNA-binding proteins.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. DESIREE;
RA   Zourelidou M.;
RL   Thesis (1997), University of East Anglia, Norwich, Norfolk,
RL   United Kingdom.
DR   EMBL; AJ314612; CAC39398.1; -.
SQ   SEQUENCE   399 AA;  44232 MW;  A118D95465799F1E CRC64;


  Query Match             13.8%;  Score 86;  DB 10;  Length 399;
  Best Local Similarity   27.3%;  Pred. No. 4.2;
  Matches   27;  Conservative   13;  Mismatches   37;  Indels   22;  Gaps    2;

Qy       2 SQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKG 61
             ||::||:   || |||:   : :|   |: |     |     | | |          ||
Db      28 SQEEEEQQSREEEGEEESGEETEEDEEPKTAHPVVKKPISQKLVQTP----------QKP 77

Qy      62 QKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTP 100
           |   :||  |            :| |    :|: :|:|
Db      78 QFSSESGSEN-----------GSGSDSEAESGNSLPSP 104


RESULT  23
Q9SKL9
ID   Q9SKL9      PRELIMINARY;      PRT;   775 AA.
AC   Q9SKL9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AT2G15190 PROTEIN.
GN   AT2G15190.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006298; AAD25573.1; -.
DR   InterPro; IPR003653; SUMO_protease.
DR   Pfam; PF02902; Ulp1_C; 1.
SQ   SEQUENCE   775 AA;  88957 MW;  CB1F20F50454A31F CRC64;


  Query Match             13.6%;  Score 85;  DB 10;  Length 775;
  Best Local Similarity   20.6%;  Pred. No. 11;
  Matches   32;  Conservative   24;  Mismatches   43;  Indels   56;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAE------EAKLKAKYPSLGQKPGGSDFLMK- 56
           :::||   |: | |:|   |:|| |  :||      |  ::      |  ||  :   :
```

```
Db     375 EEDEERKQEDEGVERQPEAEEEGGLERKAENDNESFEDSIREPNTQYGSYPGDDENTQRD 434

Qy      57 ------------------------------RLQKGQKYFDSGDYNMAKAKMKNKQL-P 83
                                          :|| :|        ::: : :||:: |
Db     435 VGDEVAEESSKDKSPTPRSSTPNFNILSEESLDVQKDKK-------RVSRGRNENKRVKP 487

Qy      84 SAGPDKNLVTGDHIPTPQ---------DLPQRKSS 109
           |   : || |   :|  :           |:| || |
Db     488 SVHAEDNLKTRKQVPRKRQKQVDSADVDVPTRKES 522


RESULT  24
Q920Q9
ID   Q920Q9      PRELIMINARY;      PRT;   967 AA.
AC   Q920Q9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SRB-RGS.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21479395; PubMed=11595167;
RA   Ikeda M., Hirokawa M., Satani N., Kinoshita T., Watanabe Y., Inoue H.,
RA   Tone S., Ishikawa T., Minatogawa Y.;
RT   "Molecular cloning and characterization of a steroid receptor-binding
RT   regulator of G-protein signaling protein cDNA.";
RL   Gene 273:207-214(2001).
DR   EMBL; AB055153; BAB63460.1; -.
SQ   SEQUENCE   967 AA;  106380 MW;  5C558885DBA63A5B CRC64;


  Query Match             13.5%;  Score 84.5;  DB 11;  Length 967;
  Best Local Similarity   28.6%;  Pred. No. 16;
  Matches   36;  Conservative   16;  Mismatches   61;  Indels   13;  Gaps    4;

Qy       3 QKQEEENPAE----ETGEEKQDTQEKEGILPERAEEAKLKAKYPSL----GQK--PGGSD 52
             |:|    | |    ||  ::::    ||  |   |    |   ||     ||:  |
Db     387 QQQLAATPTERKMFETEADEKEMPLVEGKGPGAEERTPSKDPSPSQELPPGQELPPSKDP 446

Qy      53 FLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVT 112
              : |  ||:   | | : ::     ::|||:   ||       :|  |||| || |
Db     447 SPSQELPPGQELPPSKDPSPSQELPPGQELPSS---KNPSPSQELPAGQDLPPRKESFSG 503

Qy     113 SKLAGG 118
              : | |
Db     504 QEAAPG 509


RESULT  25
Q9VLL1
ID   Q9VLL1      PRELIMINARY;      PRT;   969 AA.
AC   Q9VLL1; Q9XZU3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   D12 PROTEIN.
GN   D12 OR CG13400.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE OF 233-969 FROM N.A.
RA   Hovemann B.T., Heiermann R., Richardt A.;
RT   "Transcriptional activity at the chromosomal locus 29C.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003621; AAF52676.1; -.
DR   EMBL; AJ242723; CAB44307.1; -.
DR   FlyBase; FBgn0027490; D12.
KW   Hypothetical protein.
SQ   SEQUENCE   969 AA;  111119 MW;  6D4EAA9A51367867 CRC64;


  Query Match             13.4%;  Score 84;  DB 5;  Length 969;
  Best Local Similarity   27.4%;  Pred. No. 18;
  Matches   32;  Conservative   20;  Mismatches   51;  Indels   14;  Gaps    5;

Qy       1 MSQKQEEEN--PAEETGEEK-QDTQEKEGILP-----ERAEEAKLKAKYPSLGQKPGGSD 52
             :| : |||    ||  |: |   |:: :|:||     :|  |          : ||
Db      27 VSSQDEEETCAEAESPGQRKWTKTEQSDGMLPLSDKFQRIRELLRLEFQREISQKVEQLA 86

Qy      53 FLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPD-----KNLVTGDHIPTPQDLP 104
             : :|| :|::   |   | :     : :|:|  | |      : : || |  || ||
Db      87 EIDRRLLQGRQLLDRLRYQVVSEYYRKQQVPLTGADIAKVRGDSLFGDDIAAPQ-LP 142


RESULT  26
Q95R51
ID   Q95R51      PRELIMINARY;      PRT;   969 AA.
AC   Q95R51;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SD07884P.
GN   D12.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
```

```
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY061612; AAL29160.1; -.
SQ   SEQUENCE   969 AA;  111048 MW;  567E1A22FA17A82F CRC64;


  Query Match             13.4%;  Score 84;  DB 5;  Length 969;
  Best Local Similarity   27.4%;  Pred. No. 18;
  Matches   32;  Conservative   20;  Mismatches   51;  Indels   14;  Gaps    5;

Qy       1 MSQKQEEEN--PAEETGEEK-QDTQEKEGILP-----ERAEEAKLKAKYPSLGQKPGGSD 52
           :| : |||    ||  |: |   |:: :|:||     :|  |          : ||
Db      27 VSSQDEEETCAEAESPGQRKWTKTEQSDGMLPLSDKFQRIRELLRLEFQREISQKVEQLA 86

Qy      53 FLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPD-----KNLVTGDHIPTPQDLP 104
             : :|| :|::  |   | :     : :|:|  | |      : : || |  || ||
Db      87 EIDRRLLQGRQLLDRLRYQVVSEYYRKQQVPLTGADIAKVRGDSLFGDDIAAPQ-LP 142


RESULT  27
Q9P2P9
ID   Q9P2P9      PRELIMINARY;      PRT;  2242 AA.
AC   Q9P2P9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA1297 PROTEIN (FRAGMENT).
GN   KIAA1297.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=20181126; PubMed=10718198;
RA   Nagase T., Kikuno R., Ishikawa K., Hirosawa M., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes.XVI.
RT   The complete sequences of 150 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 7:65-73(2000).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AB037718; BAA92535.1; -.
DR   HSSP; P56276; 1TLK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 2.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00410; IG_like; 3.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 2.
KW   ATP-binding; Serine/threonine-protein kinase; Transferase.
FT   NON_TER       1      1
FT   NON_TER    2242   2242
SQ   SEQUENCE   2242 AA;  243491 MW;  497946B526A3561A CRC64;


  Query Match             13.4%;  Score 84;  DB 4;  Length 2242;
  Best Local Similarity   27.4%;  Pred. No. 47;
  Matches   29;  Conservative   16;  Mismatches   43;  Indels   18;  Gaps    4;

Qy       5 QEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKP--GGSDFLMKRLQKGQ 62
           ::|     |||       :::|  | : :||      || ||:|  | |       ::|:
Db     975 EDEALGTPETGAATPMDWQEQGRAPSQDQEAPSPEALPSPGQEPAAGASP------RRGE 1028

Qy      63 KYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKS 108
                 : :       || ||| : |  | |     :||||:|
Db    1029 ---------LRRGSSAESALPRAGP-RELGRGLHKAASVELPQRRS 1064


RESULT  28
Q9VYC3
ID   Q9VYC3      PRELIMINARY;      PRT;  281 AA.
AC   Q9VYC3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG15754 PROTEIN.
GN   CG15754.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003492; AAF48280.1; -.
DR   FlyBase; FBgn0030492; CG15754.
SQ   SEQUENCE   281 AA;  31630 MW;  34F3DD5BD11B5398 CRC64;


  Query Match             13.4%;  Score 83.5;  DB 5;  Length 281;
```

```
  Best Local Similarity   36.5%;  Pred. No. 4.8;
  Matches   27;  Conservative   12;  Mismatches   20;  Indels   15;  Gaps    4;

Qy       3 QKQEEENP-----AEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKR 57
           | ||||:|     |||  ||::| :|:|   |:       :   |:| | |||   :::
Db     120 QGQEEESPDQDGEAEEAEEEEEDEEEQEEATPQP------QPLPPALPQ-PGG---VLEE 169

Qy      58 LQKGQKYFDSGDYN 71
           |  ||   :  |||
Db     170 LAGGQVDEELEDYN 183


RESULT  29
Q9P305
ID   Q9P305      PRELIMINARY;      PRT;   131 AA.
AC   Q9P305;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   YHR132W-AP.
GN   YHR132W-A.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94378003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Keppler D., Langston Y.,
RA   Latreille P., Louis E., Macri C., Mardis E., Mouser L., Nhan M.,
RA   Rifken L., Riles L., St Peter H., Thornton L., Trevaskis E.,
RA   Vaudin M., Vaughan K., Vignati D., Wilcox L., Willis A., Wilson R.,
RA   Wohldman P., Waterston R.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Fulton L.;
RT   "The sequence of S. cerevisiae cosmid 9315.";
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Jia Y., Cherry J.M.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Cherry J.M.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U10398; AAF28366.1; -.
DR   SGD; S0007496; YHR132W-A.
SQ   SEQUENCE   131 AA;  14395 MW;  A6B6D9298D9FD013 CRC64;


  Query Match             13.3%;  Score 83;  DB 3;  Length 131;
  Best Local Similarity   32.4%;  Pred. No. 2.2;
  Matches   24;  Conservative    9;  Mismatches   29;  Indels   12;  Gaps    3;

Qy      24 KEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKM------ 77
           |||:   :    :||  |   |: |   | |  ::|  | |||||||  : || :
Db      22 KEGVDLSKLSPQELKL-YKMYGKLPSKKDLLRHKMQDRQ-YFDSGDYALKKAGVIKSDDV 79

Qy      78 ----KNKQLPSAGP 87
                :  ||   |
Db      80 IVNNSSNNLPVTNP 93


RESULT  30
Q9CUW6
ID   Q9CUW6      PRELIMINARY;      PRT;   206 AA.
AC   Q9CUW6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2900056O04RIK PROTEIN (FRAGMENT).
GN   2900056O04RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=HIPPOCAMPUS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK013708; BAB28962.1; -.
DR   MGD; MGI:1920234; 2900056O04Rik.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   InterPro; IPR000690; ZnF_matrin.
DR   InterPro; IPR003604; ZnF_U1.
DR   SMART; SM00451; ZnF_U1; 1.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
FT   NON_TER       1       1
SQ   SEQUENCE   206 AA;  23233 MW;  4316F75C83BC8F72 CRC64;


  Query Match             13.3%;  Score 83;  DB 11;  Length 206;
  Best Local Similarity   24.8%;  Pred. No. 3.7;
  Matches   35;  Conservative   16;  Mismatches   58;  Indels   32;  Gaps    4;

Qy       2 SQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPG---GSDFLMKRL 58
           | ::|:|:  ||  || :   |:|     :  |   :    |    |    | |||:  :
Db      47 SGQEEDEDDDEEEEEEGEIEAEEEFCKQVKPRETSSEQGKGSETYNPNTAYGEDFLVPVM 106

Qy      59 ----QKGQKYFDSGD------------YNMAKAKMKNKQLPSAGPDKN---------- 90
               |   |::||                |: | | ||   |   |
Db     107 GYVCQICHKFYDSNSELRFSHCKSLAHFENLQKYKAKNPTPPPTRPVSRKCAINARNALT 166

Qy      91 -LVTGDHIPTPQDLPQRKSSL 110
            | |  | |:|||  :  | :
Db     167 ALFTSSHQPSPQDTVKMPSKV 187


RESULT  31
O73793
ID   O73793      PRELIMINARY;      PRT;   487 AA.
AC   O73793;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEUROFILAMENT MEDIUM SUBUNIT.
OS   Serinus canaria (Canary).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Passeroidea;
OC   Fringillidae; Carduelinae; Serinus.
```

```
OX   NCBI_TaxID=9135;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Siepka S.M., Clayton D.F.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR   EMBL; AF053711; AAC06245.1; -.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Coiled coil; Intermediate filament.
SQ   SEQUENCE   487 AA;  53397 MW;  470F20B0892DDF43 CRC64;


  Query Match             13.3%;  Score 83;  DB 13;  Length 487;
  Best Local Similarity   25.9%;  Pred. No. 10;
  Matches   37;  Conservative   17;  Mismatches   59;  Indels   30;  Gaps    5;

Qy      3 QKQEEE---------NPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQK------ 47
          :|:|||          : ||| | ||:: :|||       |||: | |    | |
Db    143 EKEEEEAEEEEAAKSDAAEEGGSEKEEIEEKEEGEEAEEEEAEAKGKAEEAGAKVEKVKS 202

Qy     48 ---------PGGSDFL--MKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNL---VT 93
                  |  |      | :||           |:   |:  :| |:|     ||
Db    203 PPAKSPPKSPPKSPVTEQAKAVQKAAAEVGKDQKAEKAAEKAAKEEKAASPEKPATPKVT 262

Qy     94 GDHIP-TPQDLPQRKSSLVTSKL 115
               | ||:  |  : ::   |:
Db    263 SPEKPATPEKPPTPEKAITPEKV 285


RESULT  32
Q960P3
ID   Q960P3      PRELIMINARY;      PRT;   565 AA.
AC   Q960P3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LD41932P.
GN   CG1905.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Y, CN BW SP;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY051941; AAK93365.1; -.
SQ   SEQUENCE   565 AA;  63407 MW;  3F53DE0AECB8701B CRC64;


  Query Match             13.3%;  Score 83;  DB 5;  Length 565;
  Best Local Similarity   30.6%;  Pred. No. 12;
  Matches   38;  Conservative   15;  Mismatches   37;  Indels   34;  Gaps    7;

Qy      1 MSQKQEEENPAEETGEEKQDTQEKEGIL--PERAE-------EAKLKAKYPSLGQKPGGS 51
          : :|: ||    || ||:|  :|:|  |   ||||       | |||  |   | :|
Db     66 LKEKEREEKLKRET-EERQREKEREEKLKEKERAEKLKDLEKEVKLKEKEEQLKEK---- 120

Qy     52 DFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPS---------AGPDKNLV--TGDHIPTP 100
          :  :|  :| :|          | | | | | |         : | : :|  |   :|
Db    121 EKELKLKEKKEK---------DKVKEKEKSLESEKLLISATVSNPWRRVVEDTPPKLPAV 171

Qy    101 QDLP 104
          || |
Db    172 QDYP 175


RESULT  33
Q54972
ID   Q54972      PRELIMINARY;      PRT;   619 AA.
AC   Q54972;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PNEUMOCOCCAL SURFACE PROTEIN A PRECURSOR.
GN   PSPA.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yother J., Briles D.E.;
RT   "Structural and evolutionary properties of PspA, a surface protein of
RT   Streptococcus pneumoniae, as revealed by sequence analysis.";
RL   Submitted (MAR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92105030; PubMed=1729249;
RA   Yother J., Briles D.E.;
RT   "Structural properties and evolutionary relationships of PspA, a
RT   surface protein of Streptococcus pneumoniae, as revealed by sequence
RT   analysis.";
RL   J. Bacteriol. 174:601-609(1992).
DR   EMBL; M74122; AAA27018.1; -.
DR   HSSP; P00192; 1APC.
DR   InterPro; IPR002479; CW_binding.
DR   InterPro; IPR000566; Lipocln_cytFABP.
DR   Pfam; PF01473; CW_binding_1; 10.
DR   PROSITE; PS00213; LIPOCALIN; UNKNOWN_3.
KW   Signal.
FT   SIGNAL        1     31       POTENTIAL.
FT   CHAIN        32    619       PNEUMOCOCCAL SURFACE PROTEIN A.
SQ   SEQUENCE   619 AA;  68605 MW;  5AA8BDB40C2841CA CRC64;


  Query Match             13.3%;  Score 83;  DB 2;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 13;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy      8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
            | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db    354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy     66 DSGDYNMAKAKMKN 79
            : | :||   ::|
Db    414 YNTDGSMATGWLQN 427


RESULT  34
Q9U0D5
ID   Q9U0D5      PRELIMINARY;      PRT;   735 AA.
AC   Q9U0D5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KINESIN-II HOMOLOGUE.
GN   KIN1.
OS   Tetrahymena thermophila.
OC   Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC   Tetrahymenina; Tetrahymena.
OX   NCBI_TaxID=5911;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Brown J.M., Marsala C., Kosoy R., Gaertig J.;
RT   "Kinesin-II is preferentially targeted to assembling cilia and is
RT   required for ciliogenesis and normal cytokinesis in Tetrahymena.";
RL   Mol. Biol. Cell 3081:3081-3096(1999).
```

```
DR   EMBL; AJ244020; CAB51556.1; -.
DR   HSSP; P33176; 1BG2.
DR   InterPro; IPR001752; kinesin.
DR   Pfam; PF00225; kinesin; 1.
DR   PRINTS; PR00380; KINESINHEAVY.
DR   SMART; SM00129; KISc; 1.
DR   PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR   PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW   ATP-binding; Coiled coil; Microtubules; Motor protein.
SQ   SEQUENCE   735 AA;  85037 MW;  8395EDB1BB97C78D CRC64;


  Query Match             13.3%;  Score 83;  DB 5;  Length 735;
  Best Local Similarity   26.5%;  Pred. No. 16;
  Matches   31;  Conservative   21;  Mismatches   35;  Indels   30;  Gaps    4;

Qy       1 MSQKQEEE---------NPAEETGEEKQ-DTQEKEGILPERAEEAKLKAKYPSLGQKPGG 50
           : || |:|         |  ::  |:|:   :||: :| :  :: : : |     :|
Db     405 LQQKLEQEKSEVEKKYQNEVKKIEEDKKLQEKEKQQLLQQLMDKEEREQKQREQQEK--- 461

Qy      51 SDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRK 107
             |:|:||| |:    ||  | ||  | ::|  |               |:| |||
Db     462 ---LLKKLQKMQEKVIQGDEIMQKALEKERELNKA-------------QQELQQRK 501


RESULT  35
Q9SUA1
ID   Q9SUA1      PRELIMINARY;      PRT;   763 AA.
AC   Q9SUA1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYPOTHETICAL 85.2 KDA PROTEIN.
GN   T15N24.80 OR AT4G26630.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Zimmermann W., Grueneisen A., Wambutt R., Bancroft I.,
RA   Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Zimmermann W., Grueneisen A., Wambutt R., Kalicki J., Wohldmann P.,
RA   Smith A., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL078465; CAB43859.1; -.
DR   EMBL; AL161565; CAB79518.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   763 AA;  85249 MW;  D67F4F5BD4DB480B CRC64;


  Query Match             13.3%;  Score 83;  DB 10;  Length 763;
  Best Local Similarity   25.0%;  Pred. No. 17;
  Matches   31;  Conservative   25;  Mismatches   36;  Indels   32;  Gaps    6;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGI--LPERAEEAKLKAKYPSLGQKPGGSDFLMK---- 56
           :|:|||   ||  |||:: :|:|    :|:::|:   :|  ||  ::   |:   :
Db     526 EKEEEEKQEEEKAEEKEEKKEEENENGIPDKSED---EAPQPSESEEKDESEEHSEEETT 582

Qy      57 ------RLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSL 110
                 ||  |:|         :  : :||        | :|     | |: : |::||
Db     583 KKKRGSRLSAGKK--------ESAGRARNK--------KAVVAAKSSP-PEKITQKRSSA 625

Qy     111 VTSK 114
             |
Db     626 KRKK 629


RESULT  36
Q9SKM3
ID   Q9SKM3      PRELIMINARY;      PRT;   792 AA.
AC   Q9SKM3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AT2G15140 PROTEIN.
GN   AT2G15140.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006298; AAD25576.1; -.
DR   InterPro; IPR003653; SUMO_protease.
DR   Pfam; PF02902; Ulp1_C; 1.
SQ   SEQUENCE   792 AA;  91035 MW;  D39C5C03CCC28D75 CRC64;


  Query Match             13.3%;  Score 83;  DB 10;  Length 792;
  Best Local Similarity   21.3%;  Pred. No. 18;
  Matches   33;  Conservative   25;  Mismatches   41;  Indels   56;  Gaps    6;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAE------EAKLKAKYPSLGQKPGG------- 50
           :::||   |: | |:|   |:|| |  :||      |  ::      |  ||
Db     386 EEDEERKQEDEGVERQPEAEEEGGLERKAENDNESFEDSIREPNTQYGSYPGDDENTQRD 445

Qy      51 -SDFLMKR-----------------------LQKGQKYFDSGDYNMAKAKMKNKQL-P 83
             | |::                        :|| :|       ::: : :||:: |
Db     446 VGDELVEESSKDKSPTPRSSTPNFNILSEESLDVQKDKK-------RVSRGRNQNKRVKP 498

Qy      84 SAGPDKNLVTGDHIPTPQ---------DLPQRKSS 109
           |   : || |   :|  :          |:| || :
Db     499 SVHAEDNLKTRKQVPRKRQKQVDSADVDVPTRKEA 533


RESULT  37
Q9VYU0
ID   Q9VYU0      PRELIMINARY;      PRT;  2274 AA.
AC   Q9VYU0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CG1905 PROTEIN.
GN   CG1905.
OS   Drosophila melanogaster (Fruit fly).
```

```
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003487; AAF48098.1; -.
DR   FlyBase; FBgn0030337; CG1905.
SQ   SEQUENCE   2274 AA;  247009 MW;  B287311CBDF9BF20 CRC64;
```

```
  Query Match             13.3%;  Score 83;  DB 5;  Length 2274;
  Best Local Similarity   30.6%;  Pred. No. 59;
  Matches   38;  Conservative   15;  Mismatches   37;  Indels   34;  Gaps    7;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGIL--PERAE-------EAKLKAKYPSLGQKPGGS 51
           : :|: ||    || ||:|  :|:|  |   ||||       | ||| |   | :|
Db    1564 LKEKEREEKLKRET-EERQREKEREEKLKEKERAEKLKDLEKEVKLKEKEEQLKEK---- 1618

Qy      52 DFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPS---------AGPDKNLV--TGDHIPTP 100
           :  :|  :| :|          | | | | | |         : | : :|  |   :|
Db    1619 EKELKLKEKKEK---------DKVKEKEKSLESEKLLISATVSNPWRRVVEDTPPKLPAV 1669

Qy     101 QDLP 104
           || |
Db    1670 QDYP 1673


RESULT  38
Q9XGD3
ID   Q9XGD3      PRELIMINARY;      PRT;   795 AA.
AC   Q9XGD3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AT2G06430 PROTEIN.
GN   AT2G06430.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC007235; AAD26891.2; -.
DR   InterPro; IPR003653; SUMO_protease.
DR   Pfam; PF02902; Ulp1_C; 1.
SQ   SEQUENCE   795 AA;  91175 MW;  09E54A13ABF69B88 CRC64;
```

```
  Query Match             13.2%;  Score 82.5;  DB 10;  Length 795;
  Best Local Similarity   25.4%;  Pred. No. 20;
  Matches   33;  Conservative   22;  Mismatches   38;  Indels   37;  Gaps    6;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKR---LQK 60
           :::||   |: | |:|   |:|| |  :||:     | |:       : | :    :||
Db     405 EEDEERKQEDEGVERQPEAEEEGGLERKAED-----KSPTPRSTTPNFNILSEESLDVQK 459

Qy      61 GQKYFDSGDYNMAKAKMKNKQL-PSAGPDKNLVT--------------GDHIPTP---- 100
             :|        ::: : :||:: ||   : || |               ||:
Db     460 DKK-------RVSRGRNENKRVKPSVHAEDNLKTRKQAQSKKRKIIGNDGDNADNDGDND 512

Qy     101 --QDLPQRKS 108
             |  |||||
Db     513 DFQPAPQRKS 522


RESULT  39
O76719
ID   O76719      PRELIMINARY;      PRT;   335 AA.
AC   O76719;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 38.6 KDA PROTEIN.
GN   F36H12.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Blanchard M., Bradshaw H., Stellyes L.;
RT   "The sequence of C. elegans cosmid F36H12.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF078790; AAC26930.1; -.
DR   InterPro; IPR000535; MSP_domain.
DR   Pfam; PF00635; MSP_domain; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   335 AA;  38622 MW;  22FD4A1CFA8E2886 CRC64;


  Query Match           13.1%;  Score 82;  DB 5;  Length 335;
  Best Local Similarity  27.2%;  Pred. No. 8;
  Matches   41;  Conservative   20;  Mismatches   56;  Indels   34;  Gaps    7;

Qy       3 QKQEE--ENPAEETGEEKQDTQEKEGILPERAE-EAKLKAK----YPSLGQKPGGSDFLM 55
           :|:||  |   ||  |||:: ::||    |: | | |:: |     | :  :| | | :
Db     179 EKKEEKKEEKKEEKKEEKKEEEKKEEKKEEKKEDEKKVEPKPGEPKPHISIEPPG-DLMF 237

Qy      56 KRLQKGQKYFDSGDYN----MAKAKMKNKQLPSAGPDKNLV----TGDHIPTPQDLPQRK 107
           |  |: ||     : :    | | |  : |:    |    |    | |   |    | ::
Db     238 KADQQEQKKLTLTNTHDKKIMFKIKTSDNQVYLVNPVYGTVEPGKTADLTITRNKAPAKE 297

Qy     108 SSLV-----------------TSKLAGGQV 120
           : ||                 |  ||||
Db     298 AKLVIVNSVFSGDDKDLAKSFKIGKPTGGQV 328


RESULT  40
Q9VC00
ID   Q9VC00       PRELIMINARY;      PRT;  2768 AA.
AC   Q9VC00;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG13648 PROTEIN.
GN   CG13648.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003750; AAF56376.1; -.
DR   FlyBase; FBgn0039257; CG13648.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF00093; vwc; 1.
DR   SMART; SM00214; VWC; 4.
DR   PROSITE; PS01208; VWFC; 1.
SQ   SEQUENCE   2768 AA;  294032 MW;  CA929A21774E4684 CRC64;


  Query Match           13.1%;  Score 82;  DB 5;  Length 2768;
  Best Local Similarity  22.4%;  Pred. No. 92;
  Matches   32;  Conservative   24;  Mismatches   49;  Indels   38;  Gaps    5;

Qy       1 MSQKQEEENPAEETGEEK---QDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKR 57
           : : :|:: | ||: |:|   :|::||| ||     :::: :     :|   :||
Db    1227 IDESEEDKKPVEESAEDKKPVEDSEEKEKPLPTVIPASEIEKESKPEDEKKTEADFAAPT 1286

Qy      58 LQ----------KGQKYFD----------SGDYNMAKAKMKNK----QLPSAGPDKNLV 92
            |          :|  |           ||:  :  |  | :    |:| |
Db    1287 EQPEATTPAQIADTAEKEVDDKLATTSAPVSGEDELKPADEKKRTETAQIPDA------- 1339

Qy      93 TGDHIPTPQDLPQRKSSLVTSKL 115
           ||   | |:  : | |  |
Db    1340 ---EIPASTDEPESSTELPTVDL 1359


Search completed: September  7, 2002, 10:50:56
Job time: 387 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:51:16 ; Search time 15.81 Seconds
(without alignments)
296.336 Million cell updates/sec

Title: US-09-824-178-4
Perfect score: 625
Sequence: 1 MSQKQEEENPAEETGEEKQD.....DLPQRKSSLVTSKLAGGQVE 121

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	427.5	68.4	111	1	AP19_HUMAN	P56211 homo sapien
2	427.5	68.4	111	1	AP19_MOUSE	P56212 mus musculu
3	425.5	68.1	111	1	AP19_BOVIN	Q28055 bos taurus
4	90.5	14.5	168	1	YNP7_YEAST	P53897 saccharomyc
5	83.5	13.4	407	1	IE68_HSVSA	Q01042 herpesvirus
6	82	13.1	644	1	NFM_RABIT	P54938 oryctolagus
7	81.5	13.0	583	1	SHC_HUMAN	P29353 homo sapien
8	81	13.0	155	1	H2B4_VOLCA	P16868 volvox cart
9	81	13.0	325	1	ZN36_HUMAN	P17029 homo sapien
10	81	13.0	1109	1	TCF8_RAT	Q62947 rattus norv
11	79	12.6	857	1	NFM_CHICK	P16053 gallus gall
12	79	12.6	1099	1	NKX1_HUMAN	O60721 homo sapien
13	79	12.6	4967	1	RYR2_HUMAN	Q92736 homo sapien
14	78.5	12.6	260	1	DH10_ARATH	P42759 arabidopsis
15	78.5	12.6	568	1	RGS3_MOUSE	Q9dc04 mus musculu
16	78.5	12.6	1359	1	ATRX_CAEEL	Q9u7e0 caenorhabdi
17	78.5	12.6	2004	1	MOZ_HUMAN	Q92794 homo sapien
18	78	12.5	411	1	FKB3_YEAST	P38911 saccharomyc
19	77	12.3	1220	1	IF2P_HUMAN	O60841 homo sapien
20	76.5	12.2	219	1	BASP_RAT	Q05175 rattus norv
21	76.5	12.2	239	1	CALD_MELGA	P13505 meleagris g
22	76.5	12.2	771	1	CALD_CHICK	P12957 gallus gall
23	76.5	12.2	801	1	BRD2_HUMAN	P25440 homo sapien
24	76	12.2	629	1	HAP1_RAT	P54256 rattus norv
25	75.5	12.1	300	1	TRT2_RABIT	P09741 oryctolagus
26	75.5	12.1	790	1	SMY2_YEAST	P32909 saccharomyc
27	75.5	12.1	845	1	NFM_RAT	P12839 rattus norv
28	75.5	12.1	846	1	NRT_DROME	P23654 drosophila
29	75	12.0	243	1	NUKS_RAT	Q9epj0 rattus norv
30	75	12.0	357	1	FKB4_DROME	P54397 drosophila
31	75	12.0	752	1	HPR1_YEAST	P17629 saccharomyc
32	75	12.0	877	1	INCE_CHICK	P53352 gallus gall
33	75	12.0	3924	1	ANK2_HUMAN	Q01484 homo sapien
34	74.5	11.9	851	1	DD21_MOUSE	Q9jik5 mus musculu
35	74.5	11.9	1739	1	CHD2_HUMAN	O14647 homo sapien
36	74.5	11.9	4128	1	PRKD_HUMAN	P78527 homo sapien
37	73.5	11.8	1054	1	RDGB_DROME	P43125 drosophila
38	73.5	11.8	3866	1	HRX_MOUSE	P55200 mus musculu
39	73	11.7	449	1	YBB9_SCHPO	O60071 schizosacch
40	73	11.7	467	1	HEX_ADE01	Q04965 human adeno
41	73	11.7	705	1	TRDN_RABIT	Q28820 oryctolagus
42	73	11.7	807	1	YMT9_YEAST	Q04660 saccharomyc
43	73	11.7	1616	1	DNM1_HUMAN	P26358 homo sapien
44	73	11.7	2220	1	CABI_HUMAN	Q9y6j0 homo sapien
45	72.5	11.6	531	1	CALD_RAT	Q62736 rattus norv

ALIGNMENTS

```
RESULT   1
AP19_HUMAN
ID   AP19_HUMAN     STANDARD;      PRT;   111 AA.
AC   P56211; Q9UD70;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   cAMP-regulated phosphoprotein 19 (ARPP-19) [Contains: cAMP-regulated
DE   phosphoprotein 16 (ARPP-16)].
GN   ARPP19.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98318667; PubMed=9653196;
RA   Heron L., Virsolvy A., Peyrollier K., Gribble F.M., Ashcroft F.M.,
RA   Lecam A., Bataille D.;
RT   "Human alpha endosulfine, a possible regulator of sulphonylurea-
RT   sensitive KATP channel. Molecular cloning, expression and biological
RT   properties.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:8387-8391(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20219739; PubMed=10754390;
RA   Chin L.S., Singh S.K., Wang Q., Murray S.F.;
RT   "Identification of okadaic-acid-induced genes by mRNA differential
RT   display in glioma cells.";
RL   J. Biomed. Sci. 7:152-159(2000).
RN   [3]
RP   SEQUENCE OF 16-111 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=94165795; PubMed=8120638;
RA   Brene S., Lindefors N., Ehrlich M., Taubes T., Horiuchi A., Kopp J.,
RA   Hall H., Sedvall G., Greengard P., Persson H.;
RT   "Expression of mRNAs encoding ARPP-16/19, ARPP-21, and DARPP-32 in
RT   human brain tissue.";
RL   J. Neurosci. 14:985-998(1994).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS ARE PROBABLY PRODUCED BY
CC       ALTERNATIVE SPLICING.
CC   -!- PTM: SUBSTRATE FOR CAMP-DEPENDENT KINASE (BY SIMILARITY).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AJ223091; CAA11115.1; -.
DR   EMBL; AF084555; AAD52044.1; -.
```

```
DR   MIM; 605487; -.
KW   Phosphorylation; Alternative splicing; Acetylation.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    111       CAMP-REGULATED PHOSPHOPROTEIN 19.
FT   CHAIN        16    111       CAMP-REGULATED PHOSPHOPROTEIN 16.
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
FT   MOD_RES      16     16       ACETYLATION (BY SIMILARITY).
SQ   SEQUENCE   111 AA;  12192 MW;  357D3F039EDCABC1 CRC64;


  Query Match             68.4%;  Score 427.5;  DB 1;  Length 111;
  Best Local Similarity   76.4%;  Pred. No. 5.7e-30;
  Matches   84;  Conservative    9;  Mismatches   16;  Indels    1;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDS 67
           | |   : ||::: ::|    ||:|||||||:|| |||||||||| ||||||||||
Db       3 EVPEAASAEEQKEMEDKV-TSPEKAEEAKLKARYPHLGQKPGGSDFLRKRLQKGQKYFDS 61

Qy      68 GDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           |||||||||||||||:| |||  |||||||||||||||| ||| |||||
Db      62 GDYNMAKAKMKNKQLPTAAPDKTEVTGDHIPTPQDLPQRKPSLVASKLAG 111


RESULT   2
AP19_MOUSE
ID   AP19_MOUSE     STANDARD;      PRT;   111 AA.
AC   P56212;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   cAMP-regulated phosphoprotein 19 (ARPP-19).
GN   ARPP19.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,
RA   Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,
RA   Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B.,
RA   Theising B., Wylie T., Lennon G., Soares B., Wilson R., Waterston R.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- PTM: SUBSTRATE FOR CAMP-DEPENDENT KINASE (BY SIMILARITY).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AA260149; -; NOT_ANNOTATED_CDS.
DR   EMBL; W87203; -; NOT_ANNOTATED_CDS.
DR   EMBL; AA060186; -; NOT_ANNOTATED_CDS.
DR   EMBL; W70699; -; NOT_ANNOTATED_CDS.
DR   EMBL; AA041700; -; NOT_ANNOTATED_CDS.
DR   EMBL; AA500116; -; NOT_ANNOTATED_CDS.
DR   EMBL; W66839; -; NOT_ANNOTATED_CDS.
DR   MGD; MGI:1891691; Arpp19.
KW   Phosphorylation; Acetylation.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
SQ   SEQUENCE   111 AA;  12162 MW;  356089D8EEDCABC1 CRC64;


  Query Match             68.4%;  Score 427.5;  DB 1;  Length 111;
  Best Local Similarity   76.4%;  Pred. No. 5.7e-30;
  Matches   84;  Conservative    9;  Mismatches   16;  Indels    1;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDS 67
           | |   : ||::: ::|    ||:|||||||:|| |||||||||| ||||||||||
Db       3 EVPEAASAEEQKEMEDKV-TSPEKAEEAKLKARYPHLGQKPGGSDFLRKRLQKGQKYFDS 61

Qy      68 GDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           |||||||||||||||:| |||  |||||||||||||||| ||| |||||
Db      62 GDYNMAKAKMKNKQLPTAAPDKTEVTGDHIPTPQDLPQRKPSLVASKLAG 111


RESULT   3
AP19_BOVIN
ID   AP19_BOVIN     STANDARD;      PRT;   111 AA.
AC   Q28055; Q28054;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   cAMP-regulated phosphoprotein 19 (ARPP-19) [Contains: cAMP-regulated
DE   phosphoprotein 16 (ARPP-16)].
GN   ARPP19.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=90264450; PubMed=2160982;
RA   Horiuchi A., Williams K.R., Kurihara T., Nairn A.C., Greengard P.;
RT   "Purification and cDNA cloning of ARPP-16, a cAMP-regulated
RT   phosphoprotein enriched in basal ganglia, and of a related
RT   phosphoprotein, ARPP-19.";
RL   J. Biol. Chem. 265:9476-9484(1990).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS ARE PROBABLY PRODUCED BY
CC       ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: ARPP-19 FOUND IN ALL BRAIN REGIONS AND ALSO
CC       PRESENT IN NON-NEURONAL TISSUES. ARPP-16 ENRICHED IN THE CAUDATE
CC       NUCLEUS, FOUND IN LOW LEVELS IN CEREBRAL CORTEX.
CC   -!- PTM: SUBSTRATE FOR CAMP-DEPENDENT KINASE.
CC   -!- MASS SPECTROMETRY: MW=10708; MW_ERR=1; METHOD=MALDI;
CC       RANGE=16-111.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M33618; AAA30386.1; -.
DR   EMBL; M33617; AAA30385.1; -.
KW   Phosphorylation; Alternative splicing; Acetylation.
FT   INIT_MET      0      0
FT   CHAIN         1    111       CAMP-REGULATED PHOSPHOPROTEIN 19.
FT   CHAIN        16    111       CAMP-REGULATED PHOSPHOPROTEIN 16.
FT   MOD_RES       1      1       ACETYLATION.
FT   MOD_RES      16     16       ACETYLATION.
SQ   SEQUENCE   111 AA;  12222 MW;  43A64F039EDCABD5 CRC64;


  Query Match             68.1%;  Score 425.5;  DB 1;  Length 111;
  Best Local Similarity   76.4%;  Pred. No. 8.5e-30;
  Matches   84;  Conservative    9;  Mismatches   16;  Indels    1;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDS 67
           | |   : ||::: ::|    ||:|||||||:|| |||||||||| ||||||||||
Db       3 EVPEAASAEEQKEMEDKV-TSPEKAEEAKLKARYPHLGQKPGGSDFLRKRLQKGQKYFDS 61

Qy      68 GDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           |||||||||||||||:| |||  |||||||||||||||| ||| |||||
Db      62 GDYNMAKAKMKNKQLPTATPDKTEVTGDHIPTPQDLPQRKPSLVASKLAG 111
```

```
RESULT   4
YNP7_YEAST
ID   YNP7_YEAST     STANDARD;      PRT;   168 AA.
AC   P53897;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Hypothetical 18.1 kDa protein in YGP1-YCK2 intergenic region.
GN   YNL157W OR N1743.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE=96287653; PubMed=8686380;
RA   Nasr F., Becam A.-M., Herbert C.J.;
RT   "The sequence of 36.8 kb from the left arm of chromosome XIV reveals
RT   24 complete open reading frames: 18 correspond to new genes, one of
RT   which encodes a protein similar to the human myotonic dystrophy
RT   kinase.";
RL   Yeast 12:169-175(1996).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X92517; CAA63282.1; -.
DR   EMBL; Z71433; CAA96044.1; -.
DR   SGD; S0005101; YNL157W.
KW   Hypothetical protein.
FT   DOMAIN       155     160       POLY-SER.
SQ   SEQUENCE   168 AA;  18051 MW;  9534A95F1BD2CFFB CRC64;


  Query Match             14.5%;  Score 90.5;  DB 1;  Length 168;
  Best Local Similarity   31.9%;  Pred. No. 0.34;
  Matches   30;  Conservative   11;  Mismatches   30;  Indels   23;  Gaps    5;

Qy      34 EAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKN--------KQLPSA 85
           | ||   |   |: |   |     :|| :|||||||| : || ::|          ||
Db      35 EMKL---YKMYGKLPSKKDIFKHTMQK-RKYFDSGDYALQKAGIQNNDPINYGKNNLPLT 90

Qy      86 GPDKNLVTGDHIPTPQDLPQRKSSLV--TSKLAG 117
           | |           :|: :|: |    |:  ||
Db      91 NPSK---------LREDIIKRRISTCPSTASTAG 115


RESULT   5
IE68_HSVSA
ID   IE68_HSVSA     STANDARD;      PRT;   407 AA.
AC   Q01042;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   Immediate-early protein.
GN   73 OR ECLF1.
OS   Herpesvirus saimiri (strain 11).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=10383;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92333688; PubMed=1321287;
RA   Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA   Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA   Honess R.W.;
RT   "Primary structure of the herpesvirus saimiri genome.";
RL   J. Virol. 66:5047-5058(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92230228; PubMed=1314457;
RA   Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;
RT   "Analysis of nucleotide sequence of the rightmost 43 kbp of
RT   herpesvirus saimiri (HVS) L-DNA: general conservation of genetic
RT   organization between HVS and Epstein-Barr virus.";
RL   Virology 188:296-310(1992).
CC   -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-1 AND
CC       HSV-2 IE-68 (US1), EHV-1 65, EHV-4 (ORF4), PRV RSP40, AND VZV 63.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64346; CAA45696.1; -.
DR   EMBL; M86409; AAA46149.1; -.
DR   EMBL; S76368; AAB21116.1; -.
DR   PIR; G36813; EDBEQ3.
DR   PIR; S20244; S20244.
DR   HSSP; P53041; 1A17.
KW   Early protein.
FT   DOMAIN       60     241       GLU-RICH (ACIDIC).
SQ   SEQUENCE   407 AA;  46617 MW;  FFD399CA82CE136C CRC64;


  Query Match             13.4%;  Score 83.5;  DB 1;  Length 407;
  Best Local Similarity   30.6%;  Pred. No. 3.4;
  Matches   33;  Conservative   13;  Mismatches   35;  Indels   27;  Gaps    5;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAK----------YPSLGQK---- 47
           : :||   |||  || :: :| |    | ||||: : :           |  :|||
Db     203 EAEEEAEEAEEEAEEAEEAEEAEE-AEEEAEEAEEEEEEEAGPSTPRLPHYKVVGQKPSTQ 261

Qy      48 PGGSDFL--------MKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGP 87
           |||   |          || ||::   | |    | : :||    | |
Db     262 PGGVPKLCLKMQPQHRSRLPKGKQ---SHDKVPKKYQARNKFFSQAAP 306


RESULT   6
NFM_RABIT
ID   NFM_RABIT      STANDARD;      PRT;   644 AA.
AC   P54938;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neurofilament triplet M protein (160 kDa neurofilament protein)
DE   (Neurofilament medium polypeptide) (NF-M) (Fragment).
GN   NEF3 OR NEFM OR NFM.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=97055255; PubMed=8899542;
RA   Vitadello M., Vettore S., Lamar E., Chien K.R., Gorza L.;
RT   "Neurofilament M mRNA is expressed in conduction system myocytes of
RT   the developing and adult rabbit heart.";
RL   J. Mol. Cell. Cardiol. 28:1833-1844(1996).
CC   -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC       AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC   -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
CC       PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC       THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC       INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC       OF AXONAL CALIBER.
```

```
CC   -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC       OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC       LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC       COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z47378; CAA87454.1; -.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW   Glycoprotein.
FT   NON_TER       1      1
FT   DOMAIN       <1    197       ROD.
FT   DOMAIN      198    643       TAIL.
FT   DOMAIN       <1     33       COIL 1B.
FT   DOMAIN       34     50       LINKER 12.
FT   DOMAIN       51     72       COIL 2A.
FT   DOMAIN       73     76       LINKER 2.
FT   DOMAIN       77    197       COIL 2B.
FT   CARBOHYD    217    217       O-LINKED (GLCNAC) (BY SIMILARITY).
SQ   SEQUENCE   644 AA;  72450 MW;  030FDAA622889678 CRC64;


  Query Match             13.1%;  Score 82;  DB 1;  Length 644;
  Best Local Similarity   28.9%;  Pred. No. 7.6;
  Matches   22;  Conservative   17;  Mismatches   27;  Indels   10;  Gaps    2;

Qy       5 QEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKY 64
           :|||   || |:|::: :| ||:  ::|||          | :  ||         ::|:
Db     308 KEEEGEKEEEGQEEEEEEEDEGVKSDQAEEG---------GSEKEGSSKNEGEQEEGETE 358

Qy      65 FDSGDYNMAKAKMKNK 80
             : |:   |:|| : |
Db     359 AE-GEVEEAEAKEEKK 373


RESULT   7
SHC_HUMAN
ID   SHC_HUMAN      STANDARD;      PRT;   583 AA.
AC   P29353; O15290;
DT   01-DEC-1992 (Rel. 24, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   SHC transforming protein.
GN   SHC1 OR SHC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (P46SHC AND P52SHC ISOFORMS).
RX   MEDLINE=92323554; PubMed=1623525;
RA   Pelicci G., Lanfrancone L., Grignani F., McGlade J., Cavallo F.,
RA   Forni G., Nicoletti I., Grignani F., Pawson T., Pelicci P.G.;
RT   "A novel transforming protein (SHC) with an SH2 domain is implicated
RT   in mitogenic signal transduction.";
RL   Cell 70:93-104(1992).
RN   [2]
RP   SEQUENCE FROM N.A. (P66SHC ISOFORM).
RX   MEDLINE=97201514; PubMed=9049300;
RA   Migliaccio E., Mele S., Salcini A.E., Pelicci G., Lai K.M.,
RA   Superti-Furga G., Pawson T., Di Fiore P.P., Lanfrancone L.,
RA   Pelicci P.G.;
RT   "Opposite effects of the p52shc/p46shc and p66shc splicing isoforms on
RT   the EGF receptor-MAP kinase-fos signalling pathway.";
RL   EMBO J. 16:706-716(1997).
RN   [3]
RP   SEQUENCE FROM N.A. (P66SHC ISOFORM).
RA   Harun R., Smith K.K., Leek J.P., Markham A.F., Norris A.,
RA   Morrison J.F.J.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 482-583.
RX   MEDLINE=96080377; PubMed=7473762;
RA   Mikol V., Baumann G., Zurini M.G.M., Hommel U.;
RT   "Crystal structure of the SH2 domain from the adaptor protein SHC: a
RT   model for peptide binding based on X-ray and NMR data.";
RL   J. Mol. Biol. 254:86-95(1995).
RN   [5]
RP   STRUCTURE BY NMR OF 127-317.
RX   MEDLINE=96097066; PubMed=8524391;
RA   Zhou M.M., Ravichandran K.S., Olejniczak E.F., Petros A.M.,
RA   Meadows R.P., Sattler M., Harlan J.E., Wade W.S., Burakoff S.J.,
RA   Fesik S.W.;
RT   "Structure and ligand recognition of the phosphotyrosine binding
RT   domain of Shc.";
RL   Nature 378:584-592(1995).
CC   -!- FUNCTION: MAY COUPLE ACTIVATED GROWTH FACTOR RECEPTORS TO A
CC       SIGNALING PATHWAY THAT REGULATES THE PROLIFERATION OF MAMMALIAN
CC       CELLS. SHC MIGHT PARTICIPATE IN THE TRANSFORMING ACTIVITY OF
CC       ONCOGENIC TYROSINE KINASES.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- ALTERNATIVE PRODUCTS: 3 isoforms; p66shc (shown here), p52shc
CC       and p46shc; are produced by alternative initiation.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED IN MAMMALIAN TISSUES.
CC   -!- DOMAIN: THE PID DOMAIN SPECIFICALLY BINDS TO THE ASN-PRO-XAA-
CC       TYR(P) MOTIF FOUND IN MANY TYROSINE-PHOSPHORYLATED PROTEINS
CC       INCLUDING GROWTH FACTOR RECEPTORS.
CC   -!- PTM: PHOSPHORYLATED BY ACTIVATED EPIDERMAL GROWTH FACTOR
CC       RECEPTOR.
CC   -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PID DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X68148; CAA48251.1; -.
DR   EMBL; U73377; AAB49972.1; -.
DR   EMBL; Y09847; CAA70977.1; -.
DR   PIR; S25776; S25776.
DR   PDB; 1MIL; 08-NOV-96.
DR   PDB; 1SHC; 15-MAY-97.
DR   MIM; 600560; -.
DR   InterPro; IPR000050; PID_domain.
DR   InterPro; IPR000980; SH2.
DR   Pfam; PF00640; PID; 1.
DR   Pfam; PF00017; SH2; 1.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   PRINTS; PR00629; SHCPIDOMAIN.
DR   SMART; SM00462; PTB; 1.
DR   SMART; SM00252; SH2; 1.
DR   PROSITE; PS01179; PID; 1.
DR   PROSITE; PS50001; SH2; 1.
KW   SH2 domain; Phosphorylation; Growth regulation; Proto-oncogene;
KW   Alternative initiation; 3D-structure.
FT   CHAIN         1    583       SHC TRANSFORMING PROTEIN, P66SHC
FT                                ISOFORM.
FT   CHAIN       111    583       SHC TRANSFORMING PROTEIN, P52SHC
FT                                ISOFORM.
FT   CHAIN       156    583       SHC TRANSFORMING PROTEIN, P46SHC
FT                                ISOFORM.
FT   INIT_MET    111    111       FOR P52SHC ISOFORM.
```

```
FT   INIT_MET    156    156       FOR P46SHC ISOFORM.
FT   DOMAIN      156    339       PID.
FT   DOMAIN      488    579       SH2.
FT   CONFLICT      2      2       D -> N (IN REF. 3).
FT   CONFLICT     21     21       L -> M (IN REF. 3).
FT   CONFLICT     38     38       S -> P (IN REF. 3).
FT   CONFLICT     95     95       D -> V (IN REF. 3).
FT   CONFLICT    101    101       E -> D (IN REF. 3).
SQ   SEQUENCE   583 AA;  62852 MW;  8B07E8806711C933 CRC64;


  Query Match             13.0%;  Score 81.5;  DB 1;  Length 583;
  Best Local Similarity   26.2%;  Pred. No. 7.5;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   33;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|       | : :||  |  |      |: ||       :||
Db     339 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 398

Qy      47 KPGGSDFLMKRLQ-----KGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDHI 97
             ||    : |::       |:: ||   |    |: ||:   : :  |||    : |
Db     399 PVGGDPEVRKQMPPPPPCPGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS-- 453

Qy      98 PTPQDL 103
             |:||
Db     454 -APRDL 458


RESULT   8
H2B4_VOLCA
ID   H2B4_VOLCA     STANDARD;      PRT;   155 AA.
AC   P16868;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Histone H2B-IV.
OS   Volvox carteri.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Volvocaceae; Volvox.
OX   NCBI_TaxID=3067;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F. NAGARIENSIS / HK10;
RX   MEDLINE=91033024; PubMed=2227431;
RA   Mueller K., Lindauer A., Bruederlein M., Schmitt R.;
RT   "Organization and transcription of Volvox histone-encoding genes:
RT   similarities between algal and animal genes.";
RL   Gene 93:167-175(1990).
CC   -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC       H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H2B FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M31922; AAA34250.1; -.
DR   PIR; JQ0797; JQ0797.
DR   InterPro; IPR000558; Histone_H2B.
DR   InterPro; IPR000166; Histone_core.
DR   Pfam; PF00125; histone; 1.
DR   PRINTS; PR00621; HISTONEH2B.
DR   ProDom; PD000497; Histone_H2B; 1.
DR   SMART; SM00427; H2B; 1.
DR   PROSITE; PS00357; HISTONE_H2B; 1.
KW   Nuclear protein; Chromosomal protein; Nucleosome core; DNA-binding;
KW   Multigene family.
SQ   SEQUENCE   155 AA;  16841 MW;  56AC3D9C7A6F5940 CRC64;


  Query Match             13.0%;  Score 81;  DB 1;  Length 155;
  Best Local Similarity   29.2%;  Pred. No. 2;
  Matches   26;  Conservative   13;  Mismatches   42;  Indels    8;  Gaps    2;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEE---AKLKAKYPSLGQKPGGSDFLMKR 57
           |: | :|| || |  | | : : |    |:: |:   ||  || |: |    |      :
Db       1 MAPKTKEEKPASEAVEPKAEAKPKAEKAPKKKEKKAPAKKSAKEPAAGDAAEGD-----K 55

Qy      58 LQKGQKYFDSGDYNMAKAKMKNKQLPSAG 86
             :|  |   |  | :   |:  :  |  |
Db      56 KKKKAKVAKSETYKLYIYKVLKQVHPDTG 84


RESULT   9
ZN36_HUMAN
ID   ZN36_HUMAN     STANDARD;      PRT;   325 AA.
AC   P17029; P52745;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Zinc finger protein 36 (Zinc finger protein KOX18) (Fragment).
GN   ZNF36 OR KOX18 OR ZNF139.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Insulinoma;
RX   MEDLINE=96044430; PubMed=7557990;
RA   Tommerup N., Vissing H.;
RT   "Isolation and fine mapping of 16 novel human zinc finger-encoding
RT   cDNAs identify putative candidate genes for developmental and
RT   malignant disorders.";
RL   Genomics 27:259-264(1995).
RN   [2]
RP   SEQUENCE OF 167-222 FROM N.A.
RC   TISSUE=Lymphoid;
RX   MEDLINE=91145339; PubMed=2288909;
RA   Thiesen H.-J.;
RT   "Multiple genes encoding zinc finger domains are expressed in human T
RT   cells.";
RL   New Biol. 2:363-374(1990).
CC   -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U09848; AAC50264.1; -.
DR   EMBL; X52349; CAA36575.1; -.
DR   PIR; S10406; S10406.
DR   HSSP; P25490; 1ZNM.
DR   MIM; 601260; -.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00096; zf-C2H2; 6.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00355; ZnF_C2H2; 6.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 6.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 6.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Repeat.
FT   NON_TER       1      1
FT   DOMAIN      139    301       ZINC FINGERS.
```

```
FT   ZN_FING     139    161       C2H2-TYPE.
FT   ZN_FING     167    189       C2H2-TYPE.
FT   ZN_FING     195    217       C2H2-TYPE.
FT   ZN_FING     223    245       C2H2-TYPE.
FT   ZN_FING     251    273       C2H2-TYPE.
FT   ZN_FING     279    301       C2H2-TYPE.
SQ   SEQUENCE   325 AA;  36640 MW;  2659C09BD7B05708 CRC64;


  Query Match             13.0%;  Score 81;  DB 1;  Length 325;
  Best Local Similarity   21.3%;  Pred. No. 4.4;
  Matches   32;  Conservative   25;  Mismatches   47;  Indels   46;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           :||    |   | | ::  |:     | :|::  : :     ||    :|  || |:|:
Db      18 RQENYGSAFPQGGENRNENEESTSKAETSEDSASRGETTGRSQK----EFGEKRDQEGK- 72

Qy      64 YFDSGDYNMAKAKMKNKQL-----PSAGPDKNLVTGDHIP------------------TP 100
              :|:      : | ::      |: | ||  :||:  |                  ||
Db      73 ---TGERQQKNPEEKTRKEKRDSGPAIGKDKKTITGERGPREKGKGLGRSFSLSSNFTTP 129

Qy     101 QDLP--------------QRKSSLVTSKL 115
           :::|               | |||:  |:
Db     130 EEVPTGTKSHRCDECGKCFTRSSSLIRHKI 159


RESULT  10
TCF8_RAT
ID   TCF8_RAT       STANDARD;      PRT;  1109 AA.
AC   Q62947; Q62948;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Transcription factor 8 (Zinc finger homeodomain enhancer-binding
DE   protein) (Zfhep).
GN   TCF8.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 6-1109 FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=96365389; PubMed=8769566;
RA   Cabanillas A.M., Darling D.S.;
RT   "Alternative splicing gives rise to two isoforms of Zfhep, a zinc
RT   finger/homeodomain protein that binds T3-response elements.";
RL   DNA Cell Biol. 15:643-651(1996).
CC   -!- FUNCTION: ACTS AS A TRANSCRIPTIONAL REPRESSOR. BINDS TO E-BOX
CC       SEQUENCES IN THE IMMUNOGLOBULIN HEAVY CHAIN ENHANCER AS WELL AS IN
CC       THE REGULATORY REGIONS OF MANY OTHER TISSUE-SPECIFIC GENES (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1/ZFHEP-1 (SHOWN HERE) AND
CC       2/ZFHEP-2; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO DELTA-EF1/ZFH-1 FAMILY OF TWO-HANDED ZINC
CC       FINGER/HOMEODOMAIN PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U51583; AAB17130.1; -.
DR   EMBL; U51584; AAB17131.1; -.
DR   HSSP; P08046; 1A1G.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF00096; zf-C2H2; 7.
DR   SMART; SM00389; HOX; 1.
DR   SMART; SM00355; ZnF_C2H2; 7.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 6.
KW   Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
KW   Homeobox; Repressor; Activator; Metal-binding; Repeat;
KW   Alternative splicing.
FT   ZN_FING     150    173       C2H2-TYPE.
FT   ZN_FING     180    202       C2H2-TYPE.
FT   ZN_FING     220    242       C2H2-TYPE.
FT   ZN_FING     248    272       C2H2-TYPE (ATYPICAL).
FT   DNA_BIND    559    618       HOMEOBOX-LIKE.
FT   ZN_FING     881    903       C2H2-TYPE.
FT   ZN_FING     909    931       C2H2-TYPE.
FT   ZN_FING     937    958       C2H2-TYPE (ATYPICAL).
FT   DOMAIN      968   1109       GLU-RICH (ACIDIC).
FT   VARSPLIC      1    198       MISSING (IN ISOFORM 2).
SQ   SEQUENCE  1109 AA;  121626 MW;  BEFE291C8795DDA6 CRC64;


  Query Match             13.0%;  Score 81;  DB 1;  Length 1109;
  Best Local Similarity   33.3%;  Pred. No. 17;
  Matches   26;  Conservative   11;  Mismatches   21;  Indels   20;  Gaps    3;

Qy       1 MSQKQEE---ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLG------------ 45
           | : ||:   ||| ||  ||::: :|:|    | |||: :|     |
Db    1028 MEELQEDKECENPQEEEEEEEEEEEEEEEEEEEEAEEAEHEAAAAKTGGAVEEEAAQQAG 1087

Qy      46 ---QKPGGSDFLMKRLQK 60
              ||  ||:   ||| :
Db    1088 SFQQKASGSE--SKRLSE 1103


RESULT  11
NFM_CHICK
ID   NFM_CHICK      STANDARD;      PRT;   857 AA.
AC   P16053;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neurofilament triplet M protein (160 kDa neurofilament protein)
DE   (Neurofilament medium polypeptide) (NF-M).
GN   NEFM.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90174973; PubMed=2106668;
RA   Zopf D., Dineva B., Betz H., Gundelfinger E.D.;
RT   "Isolation of the chicken middle-molecular weight neurofilament
RT   (NF-M) gene and characterization of its promoter.";
RL   Nucleic Acids Res. 18:521-529(1990).
RN   [2]
RP   SEQUENCE OF 259-857 FROM N.A.
RX   MEDLINE=88112814; PubMed=3123320;
RA   Zopf D., Hermans-Borgmeyer I., Gundelfinger E.D., Betz H.;
RT   "Identification of gene products expressed in the developing chick
RT   visual system: characterization of a middle-molecular-weight
RT   neurofilament cDNA.";
RL   Genes Dev. 1:699-708(1987).
CC   -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC       AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC   -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
CC       PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC       THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC       INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC       OF AXONAL CALIBER.
CC   -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC       OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC       LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC       COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
```

```
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X17102; CAA34958.1; -.
DR   EMBL; X05558; CAA29073.1; -.
DR   PIR; A27040; A27040.
DR   PIR; S08061; S08061.
DR   PIR; S15762; S15762.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW   Glycoprotein.
FT   INIT_MET      0      0
FT   DOMAIN        1     98       HEAD.
FT   DOMAIN       99    406       ROD.
FT   DOMAIN      407    857       TAIL.
FT   DOMAIN       99    130       COIL 1A.
FT   DOMAIN      131    143       LINKER 1.
FT   DOMAIN      144    242       COIL 1B.
FT   DOMAIN      243    259       LINKER 12.
FT   DOMAIN      260    281       COIL 2A.
FT   DOMAIN      282    285       LINKER 2.
FT   DOMAIN      286    406       COIL 2B.
FT   CARBOHYD     46     46       O-LINKED (GLCNAC) (BY SIMILARITY).
FT   CARBOHYD    426    426       O-LINKED (GLCNAC) (BY SIMILARITY).
FT   CONFLICT    546    546       G -> R (IN REF. 2).
SQ   SEQUENCE   857 AA;  95704 MW;  4E2E0FC6AC64778B CRC64;


  Query Match           12.6%;  Score 79;  DB 1;  Length 857;
  Best Local Similarity  25.2%;  Pred. No. 18;
  Matches   36;  Conservative   18;  Mismatches   59;  Indels   30;  Gaps    5;

Qy       3 QKQEEE---------NPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQK------ 47
           :|:|||         : ||| | :|:: :|||       |||: | |    | |
Db     513 EKEEEEAEEEEAAKSDAAEEGGSKKEEIEEKEEGEEAEEEEAEAKGKAEEAGAKVEKVKS 572

Qy      48 ---------PGGSDFL--MKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNL---VT 93
                  |  |      | :||              |:   |:  :| |:|     ||
Db     573 PPAKSPPKSPPKSPVTEQAKAVQKAAAEVGKDQKAEKAAEKAAKEEKAASPEKPATPKVT 632

Qy      94 GDHIP-TPQDLPQRKSSLVTSKL 115
             | ||:  |  : ::   |:
Db     633 SPEKPATPEKPPTPEKAITPEKV 655


RESULT  12
NKX1_HUMAN
ID   NKX1_HUMAN     STANDARD;      PRT;  1099 AA.
AC   O60721; O43485; O75184;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Sodium/potassium/calcium exchanger 1 precursor (Na(+)/K(+)/Ca(2+)-
DE   exchange protein 1) (Retinal rod Na-Ca+K exchanger).
GN   SLC24A1 OR NCKX1 OR KIAA0702.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Retina;
RX   MEDLINE=99072302; PubMed=9856482;
RA   Tucker J.E., Winkfein R.J., Murthy S.K., Friedman J.S., Walter M.A.,
RA   Demetrick D.J., Schnetkamp P.P.M.;
RT   "Chromosomal localization and genomic organization of the human
RT   retinal rod Na-Ca+K exchanger.";
RL   Hum. Genet. 103:411-414(1998).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Retina;
RX   MEDLINE=98138491; PubMed=9478004;
RA   Tucker J.E., Winkfein R.J., Cooper C.B., Schnetkamp P.P.M.;
RT   "cDNA cloning of the human retinal rod Na-Ca+K exchanger: comparison
RT   with a revised bovine sequence.";
RL   Invest. Ophthalmol. Vis. Sci. 39:435-440(1998).
RN   [3]
RP   SEQUENCE OF 1-999 FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811;
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
CC   -!- FUNCTION: Critical component of the visual transduction cascade,
CC       controlling the calcium concentration of outer segments during
CC       light and darkness. Light causes a rapid lowering of cytosolic
CC       free calcium in the outer segment of both retinal rod and cone
CC       photoreceptors and the light-induced lowering of calcium is caused
CC       by extrusion via this protein which plays a key role in the
CC       process of light adaptation. Transports one Ca(2+) and one K(+) in
CC       exchange for four Na(+).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
CC       produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Found only in the outer segments of retinal
CC       rod photoreceptors.
CC   -!- SIMILARITY: BELONGS TO THE SLC24A FAMILY OF TRANSPORTERS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF062921; AAC16732.1; -.
DR   EMBL; AF062922; AAC77912.1; -.
DR   EMBL; AF026132; AAB97832.1; -.
DR   EMBL; AB014602; BAA31677.1; ALT_SEQ.
DR   MIM; 603617; -.
DR   InterPro; IPR002613; Na_Ca_Ex.
DR   Pfam; PF01699; Na_Ca_Ex; 2.
KW   Vision; Transport; Symport; Transmembrane; Glycoprotein;
KW   Phosphorylation; Signal; Calcium transport; Repeat;
KW   Alternative splicing.
FT   SIGNAL        1     38       POTENTIAL.
FT   CHAIN        39   1099       SODIUM/POTASSIUM/CALCIUM EXCHANGER 1.
FT   DOMAIN       39    452       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    453    473       POTENTIAL.
FT   DOMAIN      474    497       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    498    518       POTENTIAL.
FT   DOMAIN      519    522       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    523    543       POTENTIAL.
FT   DOMAIN      544    563       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    564    584       POTENTIAL.
FT   DOMAIN      585    585       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    586    606       POTENTIAL.
FT   DOMAIN      607    907       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    908    928       POTENTIAL.
FT   DOMAIN      929    935       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    936    956       POTENTIAL.
FT   DOMAIN      957    971       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    972    992       POTENTIAL.
```

```
FT   DOMAIN      993   1010       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1011   1031       POTENTIAL.
FT   DOMAIN     1032   1039       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1040   1060       POTENTIAL.
FT   DOMAIN     1061   1068       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1069   1089       POTENTIAL.
FT   DOMAIN     1090   1099       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      494    534       ALPHA-1.
FT   REPEAT      979   1010       ALPHA-2.
FT   DOMAIN      860    890       POLY-GLU.
FT   MOD_RES     658    658       PHOSPHORYLATION (POTENTIAL).
FT   CARBOHYD    290    290       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    631    648       MISSING (IN ISOFORM 2).
FT   CONFLICT    516    516       V -> I (IN REF. 2).
SQ   SEQUENCE   1099 AA;  121374 MW;  3911856BB088B5FD CRC64;


  Query Match             12.6%;  Score 79;  DB 1;  Length 1099;
  Best Local Similarity   28.0%;  Pred. No. 24;
  Matches   21;  Conservative   14;  Mismatches   28;  Indels   12;  Gaps    1;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMK------ 56
           :::|||   ||  ||::: :|:|    |:  |  |   :|   ||     ||:
Db     863 EEEEEEQEEEEEEEEQEEEEEEEEEEEEKGNEEPLSLDWPETRQKQAIYLFLLPIVFPLW 922

Qy      57 ------RLQKGQKYF 65
                 | |: :|:|
Db     923 LTVPDVRRQESRKFF 937


RESULT  13
RYR2_HUMAN
ID   RYR2_HUMAN     STANDARD;      PRT;  4967 AA.
AC   Q92736; Q15411;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Ryanodine receptor 2 (Cardiac muscle-type ryanodine receptor) (RyR2)
DE   (RYR-2) (Cardiac muscle ryanodine receptor-calcium release channel)
DE   (hRYR-2).
GN   RYR2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Heart muscle;
RX   MEDLINE=96404895; PubMed=8809036;
RA   Tunwell R.E.A., Wickenden C., Bertrand B.M.A., Shevchenko V.I.,
RA   Walsh M.B., Allen P.D., Lai F.A.;
RT   "The human cardiac muscle ryanodine receptor-calcium release channel:
RT   identification, primary structure and topological analysis.";
RL   Biochem. J. 318:477-487(1996).
RN   [2]
RP   SEQUENCE OF 9-87 AND 533-681 FROM N.A., DEVELOPMENTAL STAGE, AND
RP   INDUCTION.
RC   TISSUE=Myometrium, and Heart muscle;
RX   MEDLINE=97220386; PubMed=9148749;
RA   Awad S.S., Lamb H.K., Morgan J.M., Dunlop W., Gillespie J.I.;
RT   "Differential expression of ryanodine receptor RyR2 mRNA in the non-
RT   pregnant and pregnant human myometrium.";
RL   Biochem. J. 322:777-783(1997).
RN   [3]
RP   SEQUENCE OF 4292-4479 FROM N.A., AND TISSUE SPECIFICITY.
RC   TISSUE=Cerebellum, and Hippocampus;
RX   MEDLINE=98268728; PubMed=9607712;
RA   Martin C., Chapman K.E., Seckl J.R., Ashley R.H.;
RT   "Partial cloning and differential expression of ryanodine
RT   receptor/calcium-release channel genes in human tissues including the
RT   hippocampus and cerebellum.";
RL   Neuroscience 85:205-216(1998).
RN   [4]
RP   VARIANTS VTSIP L-2246; S-2474; K-4104 AND C-4497.
RX   PubMed=11208676;
RA   Priori S.G., Napolitano C., Tiso N., Memmi M., Vignati G., Bloise R.,
RA   Sorrentino V.V., Danieli G.A.;
RT   "Mutations in the cardiac ryanodine receptor gene (hRyR2) underlie
RT   catecholaminergic polymorphic ventricular tachycardia.";
RL   Circulation 103:196-200(2001).
RN   [5]
RP   VARIANTS VPFT S-2328; R-4201 AND F-4653, AND VARIANT R-2958.
RX   MEDLINE=21112862; PubMed=11157710;
RA   Laitinen P.J., Brown K.M., Piippo K., Swan H., Devaney J.M.,
RA   Brahmbhatt B., Donarum E.A., Marino M., Tiso N., Viitasalo M.,
RA   Toivonen L., Stephan D.A., Kontula K.;
RT   "Mutations of the cardiac ryanodine receptor (RyR2) gene in familial
RT   polymorphic ventricular tachycardia.";
RL   Circulation 103:485-490(2001).
RN   [6]
RP   VARIANTS ARVD2 PRO-433 AND ILE-2386.
RX   MEDLINE=21096894; PubMed=11159936;
RA   Tiso N., Stephan D.A., Nava A., Bagattin A., Devaney J.M., Stanchi F.,
RA   Larderet G., Brahmbhatt B., Brown K., Bauce B., Muriago M., Basso C.,
RA   Thiene G., Danieli G.A., Rampazzo A.;
RT   "Identification of mutations in the cardiac ryanodine receptor gene in
RT   families affected with arrhythmogenic right ventricular
RT   cardiomyopathy type 2 (ARVD2).";
RL   Hum. Mol. Genet. 10:189-194(2001).
CC   -!- FUNCTION: Communication between transverse-tubules and
CC       sarcoplasmic reticulum. Contraction of cardiac muscle is triggered
CC       by release of calcium ions from SR following depolarization of T-
CC       tubules (By similarity).
CC   -!- SUBUNIT: Homotetramer (Potential).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
CC       produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Heart muscle, brain (cerebellum and
CC       hippocampus) and placenta.
CC   -!- DEVELOPMENTAL STAGE: Expressed in myometrium during pregnancy.
CC   -!- INDUCTION: By TGF-beta.
CC   -!- DISEASE: Defects in RYR2 are a cause of arrhythmogenic right
CC       ventricular cardiomyopathy type 2 (ARVD2). ARVD2 is an autosomal
CC       dominant disease. Characterized by partial degeneration of the
CC       myocardium of the right ventricle, electrical instability and
CC       sudden death. It is clinically defined by electrocardiographic and
CC       angiographic criteria; pathologic findings, replacement of
CC       ventricular myocardium with fatty and fibrous elements,
CC       preferentially involve the right ventricular free wall.
CC   -!- DISEASE: Defects in RYR2 are a cause of stress-induced polymorphic
CC       ventricular tachycardia (VTSIP) also known as catecholaminergic
CC       polymorphic ventricular tachycardia. VTSIP is genetic
CC       arrhythmogenic disorder characterized by stress-induced,
CC       bidirectional ventricular tachycardia that may degenerate into
CC       cardiac arrest and cause sudden death.
CC   -!- DISEASE: Defects in RYR2 are a cause of familial polymorphic
CC       ventricular tachycardia (VPFT). VPFT is an autosomal-dominant,
CC       inherited disease with a relatively early onset and a mortality
CC       rate of approximately 30% by the age of 30 years. Phenotypically,
CC       it is characterized by salvoes of bidirectional and polymorphic
CC       ventricular tachycardias in response to vigorous exercise, with no
CC       structural evidence of myocardial disease.
CC   -!- MISCELLANEOUS: The calcium release channel is modulated by calcium
CC       ions, magnesium ions, ATP and calmodulin.
CC   -!- MISCELLANEOUS: The calcium release channel activity resides in the
CC       C-terminal region while the remaining part of the protein
CC       constitutes the 'foot' structure spanning the junctional gap
CC       between the SR and the T-tubule. It is possible that the foot
CC       structure interacts with the cytoplasmic region of the
CC       dihydropyridine receptor.
CC   -!- MISCELLANEOUS: Ryanodine is an alkaloid that binds to the Ca-
CC       release channel in junctional SR and modulates its activity.
CC   -!- SIMILARITY: BELONGS TO THE RYANODINE RECEPTOR FAMILY.
CC   -!- SIMILARITY: CONTAINS 3 SPRY DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X98330; CAA66975.1; -.
DR   EMBL; Y08218; CAA69395.1; -.
DR   EMBL; X91869; CAA62975.1; -.
DR   EMBL; AJ002511; CAA05502.1; -.
DR   MIM; 180902; -.
DR   MIM; 192605; -.
DR   MIM; 600996; -.
DR   MIM; 604772; -.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001682; Channel_pore_Ca_Na.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR003608; MIR.
DR   InterPro; IPR000699; RYDR_ITPR.
DR   InterPro; IPR003032; RyR.
DR   InterPro; IPR001215; Ryanodn_receptor.
DR   InterPro; IPR003877; SPRY.
DR   InterPro; IPR003878; SPRY_domain.
DR   Pfam; PF00036; efhand; 2.
DR   Pfam; PF00520; ion_trans; 1.
DR   Pfam; PF02815; MIR; 4.
DR   Pfam; PF01365; RYDR_ITPR; 2.
DR   Pfam; PF02026; RyR; 4.
DR   Pfam; PF00622; SPRY; 3.
DR   PRINTS; PR00795; RYANODINER.
DR   SMART; SM00472; MIR; 4.
DR   SMART; SM00449; SPRY; 3.
KW   Receptor; Transmembrane; Ionic channel; Calcium channel; Repeat;
KW   Phosphorylation; Glycoprotein; Multigene family; Calmodulin-binding;
KW   Alternative splicing; Disease mutation; Polymorphism.
FT   DOMAIN       1   3089       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM  3090   3109       M' (POTENTIAL).
FT   TRANSMEM  3153   3171       M'' (POTENTIAL).
FT   TRANSMEM  3940   3959       M1 (POTENTIAL).
FT   TRANSMEM  3978   3995       M2 (POTENTIAL).
FT   TRANSMEM  4233   4256       M3 (POTENTIAL).
FT   TRANSMEM  4294   4314       M4 (POTENTIAL).
FT   TRANSMEM  4500   4520       M5 (POTENTIAL).
FT   TRANSMEM  4578   4600       M6 (POTENTIAL).
FT   TRANSMEM  4720   4740       M7 (POTENTIAL).
FT   TRANSMEM  4768   4786       M8 (POTENTIAL).
FT   TRANSMEM  4810   4827       M9 (POTENTIAL).
FT   TRANSMEM  4845   4867       M10 (POTENTIAL).
FT   DOMAIN     670    808       SPRY 1.
FT   DOMAIN    1098   1221       SPRY 2.
FT   DOMAIN    1423   1561       SPRY 3.
FT   DOMAIN    4414   4455       GLU-RICH (ACIDIC).
FT   DOMAIN     853   2925       4 X APPROXIMATE REPEATS.
FT   REPEAT     853    966       1.
FT   REPEAT     967   1080       2.
FT   REPEAT    2692   2810       3.
FT   REPEAT    2812   2925       4.
FT   BINDING   2618   3015       MODULATOR (POTENTIAL).
FT   BINDING   2774   2806       CALMODULIN (POTENTIAL).
FT   BINDING   2876   2897       CALMODULIN (POTENTIAL).
FT   BINDING   2997   3015       CALMODULIN (POTENTIAL).
FT   MOD_RES   2808   2808       PHOSPHORYLATION (BY CAM-KINASE) (BY
FT                               SIMILARITY).
FT   CARBOHYD  3427   3427       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  3536   3536       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  3770   3770       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  3864   3864       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  3865   3865       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  3905   3905       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4075   4075       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4104   4104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4111   4111       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4794   4794       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC  3715   3715       E -> EVTGSQRSK (IN ISOFORM 2).
FT   VARIANT    433    433       L -> P (IN ARDV2).
FT                               /FTId=VAR_011395.
FT   VARIANT   2246   2246       S -> L (IN VTSIP).
FT                               /FTId=VAR_011396.
FT   VARIANT   2328   2328       P -> S (IN FPVT).
FT                               /FTId=VAR_011397.
FT   VARIANT   2386   2386       N -> I (IN ARDV2).
FT                               /FTId=VAR_011398.
FT   VARIANT   2474   2474       R -> S (IN VTSIP).
FT                               /FTId=VAR_011399.
FT   VARIANT   2958   2958       Q -> R.

  Query Match             12.6%;  Score 79;  DB 1;  Length 4967;
  Best Local Similarity   23.9%;  Pred. No. 1.2e+02;
  Matches   28;  Conservative   17;  Mismatches   42;  Indels   30;  Gaps    3;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL----GQKPGGSDFLMKRL 58
           |:||:   |||  | :  : ::|   |:|:| | | |   |     |:         |::
Db    4425 KEEEKEEKEETKSEPEKAEGEDGEKEEKAKEDKGKQKLRQLHTHRYGEPEVPESAFWKKI 4484

Qy      59 ----QKGQKYFDSGDYNM--------------------AKAKMKNKQLPSAGPDKN 90
               ||   ||    |||                    : :: |:||:    :|
Db    4485 IAYQQKLLNYFARNFYNMRMLALFVAFAINFILLFYKVSTSSVVEGKELPTRSSSEN 4541


RESULT  14
DH10_ARATH
ID   DH10_ARATH     STANDARD;      PRT;   260 AA.
AC   P42759;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Dehydrin ERD10 (Low-temperature-induced protein LTI45).
GN   ERD10 OR LTI45 OR LTI29 OR AT1G20450 OR F5M15.21.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=94348723; PubMed=8069491;
RA   Kiyosue T., Yamaguchi-Shinozaki K., Shinozaki K.;
RT   "Characterization of two cDNAs (ERD10 and ERD14) corresponding to
RT   genes that respond rapidly to dehydration stress in Arabidopsis
RT   thaliana.";
RL   Plant Cell Physiol. 35:225-231(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. LANDSBERG ERECTA; TISSUE=Leaf;
RX   MEDLINE=96046757; PubMed=7579189;
RA   Welin B.V., Olson A., Palva E.T.;
RT   "Structure and organization of two closely related low-temperature-
RT   induced dhn/lea/rab-like genes in Arabidopsis thaliana L. Heynh.";
RL   Plant Mol. Biol. 29:391-395(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
```

```
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
RN   [4]
RP   SEQUENCE OF 56-260 FROM N.A.
RC   STRAIN=CV. LANDSBERG ERECTA; TISSUE=Leaf;
RX   MEDLINE=95035985; PubMed=7948863;
RA   Welin B.V., Olson A., Nylander M., Palva E.T.;
RT   "Characterization and differential expression of dhn/lea/rab-like
RT   genes during cold acclimation and drought stress in Arabidopsis
RT   thaliana.";
RL   Plant Mol. Biol. 26:131-144(1994).
CC   -!- TISSUE SPECIFICITY: IN STEMS, CAULINE LEAVES, ROOTS AND FLOWERS.
CC       LOW LEVELS FOUND IN MATURING SEEDS. ABSENT IN DRY SEEDS.
CC   -!- INDUCTION: BY DEHYDRATION, COLD STRESS AND ABSCISIC ACID (ABA).
CC       INDUCTION BY DEHYDRATION OCCURS AFTER 1 H AND INCREASES TO A
CC       MAXIMUM AFTER 10 H. COLD STRESS INDUCTION PEAKS AT 1 H AND 5 H
CC       AFTER START OF COLD EXPOSURE.
CC   -!- SIMILARITY: BELONGS TO THE PLANT DEHYDRIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D17714; BAA04568.1; -.
DR   EMBL; X90958; CAA62448.1; -.
DR   EMBL; AC027665; AAF79613.1; -.
DR   EMBL; X77614; CAA54705.1; -.
DR   InterPro; IPR000167; Dehydrin.
DR   Pfam; PF00257; dehydrin; 1.
DR   PROSITE; PS00315; DEHYDRIN_1; 1.
DR   PROSITE; PS00823; DEHYDRIN_2; 2.
KW   Dehydrin; Repeat; Multigene family.
FT   DOMAIN       71     74       POLY-GLU.
FT   DOMAIN       88     92       POLY-GLU.
FT   DOMAIN      104    112       POLY-SER.
FT   DOMAIN      121    129       POLY-LYS.
FT   DOMAIN      183    186       POLY-GLU.
FT   DOMAIN      184    247       2 X 21 AA REPEATS, LYS-RICH.
FT   REPEAT      184    204       1.
FT   REPEAT      227    247       2.
FT   CONFLICT     56     56       H -> G (IN REF. 3).
FT   CONFLICT    144    144       Q -> E (IN REF. 3).
SQ   SEQUENCE   260 AA;  29548 MW;  318B6FC6B4A1AE5A CRC64;


  Query Match           12.6%;  Score 78.5;  DB 1;  Length 260;
  Best Local Similarity  28.4%;  Pred. No. 5.6;
  Matches   42;  Conservative   19;  Mismatches   42;  Indels   45;  Gaps   11;

Qy       2 SQKQEEENPAEETGEEKQDTQEK---EGILPERAEEA------KLKAKYPSLGQKPGGSD 52
           |   :||    | ||:|:  ::|   ||   :  ||       ::| |:| ||:|||| |
Db     109 SSSSDEEG---EDGEKKKKEKKKKIVEGDHVKTVEEENQGVMDRIKEKFP-LGEKPGGDD 164

Qy      53 FLM------------KRLQKGQKYFDSGDYNMAKAKMKNKQLP--SAGPDKNLVTGD-- 95
             :             |  :: :| |      | | | |   ||  |  |: : |
Db     165 VPVVTTMPAPHSVEDHKPEEEEKKGF------MDKIKEK---LPGHSKKPEDSQVVNTTP 215

Qy      96 --HIPTP-QDLPQRKSSL---VTSKLAG 117
               ||  |:|: |      :  || |
Db     216 LVETATPIADIPEEKKGFMDKIKEKLPG 243


RESULT  15
```

```
RGS3_MOUSE
ID   RGS3_MOUSE     STANDARD;      PRT;   568 AA.
AC   Q9DC04; Q9JL22; Q9JL23;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Regulator of G-protein signaling 3 (RGS3).
GN   RGS3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (LONG ISOFORM).
RC   STRAIN=C57BL/6J; TISSUE=Lung;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [2]
RP   PARTIAL SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=20243574; PubMed=10779778;
RA   Reif K., Cyster J.G.;
RT   "RGS molecule expression in murine B lymphocytes and ability to
RT   down-regulate chemotaxis to lymphoid chemokines.";
RL   J. Immunol. 164:4720-4729(2000).
CC   -!- FUNCTION: INHIBITS SIGNAL TRANSDUCTION BY INCREASING THE GTPASE
CC       ACTIVITY OF G PROTEIN ALPHA SUBUNITS THEREBY DRIVING THEM INTO
CC       THEIR INACTIVE GDP-BOUND FORM.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: PHOSPHORYLATED BY CYCLIC GMP-DEPENDENT PROTEIN KINASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 RGS DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AK004648; BAB23439.1; -.
DR   EMBL; AF215670; AAF34627.1; -.
DR   EMBL; AF215669; AAF34626.1; -.
DR   MGD; MGI:1354734; Rgs3.
DR   InterPro; IPR000342; RGS.
DR   Pfam; PF00615; RGS; 1.
DR   PRINTS; PR01301; RGSPROTEIN.
DR   ProDom; PD001580; RGS; 1.
DR   SMART; SM00315; RGS; 1.
DR   PROSITE; PS50132; RGS; 1.
KW   Signal transduction inhibitor; Alternative splicing; Phosphorylation.
FT   DOMAIN      443    559       RGS.
```

```
FT   VARSPLIC      1    375       MISSING (IN SHORT ISOFORM).
FT   VARSPLIC    376    397       KMSGTDLTECDEASRKRKSKNI -> MLRGMYLTRNGNLQR
FT                                RHTMKE (IN SHORT ISOFORM).
SQ   SEQUENCE   568 AA;  61558 MW;  91939E22F3DA9F5C CRC64;


  Query Match             12.6%;  Score 78.5;  DB 1;  Length 568;
  Best Local Similarity   30.5%;  Pred. No. 13;
  Matches   32;  Conservative   13;  Mismatches   47;  Indels   13;  Gaps    4;

Qy      13 ETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQK-PGGSDF-------LMKRLQKGQKY 64
           ||  ::::    ||  |   | |   :| || ||: | | |          : |  ||
Db       3 ETEADEKEMPLVEGKGPGAEEPA--PSKNPSPGQELPPGQDLPPSKDPSPSQELPAGQDL 60

Qy      65 FDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSS 109
             | | : ::     : ||   | |:      :|  |||| || |
Db      61 PPSKDPSPSQELPAGQDLP---PSKDPSPSQELPVGQDLPPRKDS 102


RESULT  16
ATRX_CAEEL
ID   ATRX_CAEEL     STANDARD;      PRT;  1359 AA.
AC   Q9U7E0; O02061;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
GN   XNP-1 OR B0041.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=99365296; PubMed=10433961;
RA   Villard L., Fontes M., Ewbank J.J.;
RT   "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RT   the human XNP/ATR-X gene.";
RL   Gene 236:13-19(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Fulton R., Wohldmann P.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC       GENE EXPRESSION BY AFFECTING CHROMATIN (Potential).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF134186; AAD55361.1; -.
DR   EMBL; AF000196; AAC24256.1; -.
DR   WormPep; B0041.7; CE17314.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
KW   DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding.
FT   NP_BIND     496    503       ATP (POTENTIAL).
FT   SITE        636    639       DEAH BOX.
FT   DOMAIN       67     70       POLY-ASP.
FT   DOMAIN      266    272       POLY-GLU.
FT   DOMAIN      276    281       POLY-LYS.
FT   DOMAIN      372    375       POLY-LYS.
FT   DOMAIN      603    608       POLY-LYS.
FT   DOMAIN      859    862       POLY-LYS.
FT   CONFLICT    479    479       C -> F (IN REF. 2).
SQ   SEQUENCE   1359 AA;  156191 MW;  EB4342547D4F4E64 CRC64;


  Query Match             12.6%;  Score 78.5;  DB 1;  Length 1359;
  Best Local Similarity   26.9%;  Pred. No. 34;
  Matches   32;  Conservative   24;  Mismatches   48;  Indels   15;  Gaps    5;

Qy       2 SQKQEEENPAEETGEE---KQDTQEKEGILPERAE--EAKLKAKYPSLGQKPGGSDFLMK 56
           ::||     :||: ||   |:  : ||  : :|||  |   : : ||   | |     :|
Db     147 TKKQTSSESSEESEEERKVKKSKKNKEKSVKKRAETSEESDEDEKPSKKSKKG-----LK 201

Qy      57 RLQKGQKYFDSGD-YNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSK 114
           :  | :   :| |   : |:| |:|::      |   : |  | :   :|| |  :|:
Db     202 KKAKSESESESEDEKEVKKSKKKSKKVVK----KESESEDEAPEKKKTEKRKRSKTSSE 256


RESULT  17
MOZ_HUMAN
ID   MOZ_HUMAN      STANDARD;      PRT;  2004 AA.
AC   Q92794;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Monocytic leukemia zinc finger protein (Zinc finger protein 220).
GN   ZNF220 OR MOZ.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96376968; PubMed=8782817;
RA   Borrow J., Stanton V.P. Jr., Andresen J.M., Becher R., Behm F.G.,
RA   Chaganti R.S.K., Civin C.I., Disteche C., Dube I., Frischauf A.M.,
RA   Horsman D., Mitelman F., Volinia S., Watmore A.E., Housman D.E.;
RT   "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT   a putative acetyltransferase to the CREB-binding protein.";
RL   Nat. Genet. 14:33-41(1996).
CC   -!- FUNCTION: MAY REPRESENT A CHROMATIN-ASSOCIATED ACETYLTRANSFERASE.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DISEASE: PARTICIPATES IN A T(8;16)(P11;P13) CHROMOSOMAL
CC       TRANSLOCATION THAT PRODUCES A MOZ-CBP CHIMAERA OBSERVED IN THE
CC       M4/M5 SUBTYPE OF ACUTE MYELOID LEUKEMIA.
CC   -!- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
CC   -!- SIMILARITY: BELONGS TO THE MYST (SAS/MOZ) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U47742; AAC50662.1; -.
DR   HSSP; Q60631; 4GBQ.
DR   MIM; 601408; -.
DR   InterPro; IPR001386; Linker_histone.
DR   InterPro; IPR002717; MOZ_SAS.
DR   InterPro; IPR001965; PHD.
DR   Pfam; PF01853; MOZ_SAS; 1.
DR   Pfam; PF00628; PHD; 2.
DR   SMART; SM00526; H15; 1.
DR   SMART; SM00249; PHD; 2.
KW   Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;
KW   Nuclear protein.
FT   ZN_FING     206    256       PHD-TYPE 1.
FT   ZN_FING     259    313       PHD-TYPE 2.
FT   DOMAIN      371    379       POLY-SER.
```

```
FT   ZN_FING      538    560       C2HC-TYPE.
FT   DOMAIN       788    801       POLY-GLU.
FT   DOMAIN       989    995       POLY-GLU.
FT   DOMAIN      1019   1026       POLY-ARG.
FT   DOMAIN      1069   1078       POLY-GLU.
FT   DOMAIN      1147   1150       POLY-LYS.
FT   DOMAIN      1221   1242       GLU-RICH.
FT   DOMAIN      1267   1302       GLU-RICH.
FT   DOMAIN      1411   1414       POLY-GLU.
FT   DOMAIN      1593   1597       POLY-SER.
FT   DOMAIN      1643   1704       GLN/PRO-RICH.
FT   DOMAIN      1897   1977       MET-RICH.
FT   SITE        1546   1547       BREAKPOINT FOR TRANSLOCATION TO FORM
FT                                 MOZ-CBP.
SQ   SEQUENCE   2004 AA;  225054 MW;  9FFBBAC3792854BA CRC64;


  Query Match             12.6%;  Score 78.5;  DB 1;  Length 2004;
  Best Local Similarity   20.4%;  Pred. No. 51;
  Matches   28;  Conservative   29;  Mismatches   45;  Indels   35;  Gaps    4;

Qy       4 KQEEENP--AEETGEEKQDTQEKEGILPERAEEAKLK------------------AKYP 42
           ::||| |  :||  : ::: || |   ||  |:|  :                    |
Db    1271 EEEEEKPRVSEEQRQSEEEQQELEEPEPEEEEDAAAETAQNDDHDADDEDDGHLESTKKK 1330

Qy      43 SLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQD 102
             | ::|   |  :|     |: |   :   :: |:|:|:                : : ::
Db    1331 ELEEQPTRED--VKEEPGVQESFLDANMQKSREKIKDKE-----------ETELDSEEE 1376

Qy     103 LPQRKSSLVTSKLAGGQ 119
             |   :|:|: ::|| :
Db    1377 QPSHDTSVVSEQMAGSE 1393


RESULT  18
FKB3_YEAST
ID   FKB3_YEAST     STANDARD;      PRT;   411 AA.
AC   P38911;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   FK506-binding nuclear protein (Peptidyl-prolyl cis-trans isomerase)
DE   (PPiase) (EC 5.2.1.8) (Proline rotamase) (Nucleolar proline
DE   isomerase) (FKBP-70).
GN   FPR3 OR NPI46 OR YML074C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YNN 214;
RX   MEDLINE=95050937; PubMed=7525596;
RA   Benton B.M., Zang J.-H., Thorner J.;
RT   "A novel FK506- and rapamycin-binding protein (FPR3 gene product) in
RT   the yeast Saccharomyces cerevisiae is a proline rotamase localized to
RT   the nucleolus.";
RL   J. Cell Biol. 127:623-639(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=94327703; PubMed=8051210;
RA   Shan X., Xue Z., Melese T.;
RT   "Yeast NPI46 encodes a novel prolyl cis-trans isomerase that is
RT   located in the nucleolus.";
RL   J. Cell Biol. 126:853-862(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95010686; PubMed=7925954;
RA   Manning-Krieg U.C., Henriquez R., Cammas F., Graff P.,
RA   Gaveriaux S., Movva N.R.;
RT   "Purification of FKBP-70, a novel immunophilin from Saccharomyces
RT   cerevisiae, and cloning of its structural gene, FPR3.";
RL   FEBS Lett. 352:98-103(1994).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Brown D., Bowman S., Barrell B.G., Rajandream M.A.;
RL   Submitted (OCT-1994) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   PHOSPHORYLATION OF TYR-184 AND SER-186.
RX   MEDLINE=97294697; PubMed=9148902;
RA   Wilson L.K., Dhillon N., Thorner J., Martin G.S.;
RT   "Casein kinase II catalyzes tyrosine phosphorylation of the yeast
RT   nucleolar immunophilin Fpr3.";
RL   J. Biol. Chem. 272:12961-12967(1997).
CC   -!- FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS. FK506- AND
CC       RAPAMYCIN-BINDING PROTEIN. SPECIFICALLY BINDS NUCLEAR LOCALIZATION
CC       SEQUENCES. MAY BE INVOLVED IN THE ASSEMBLY OR FOLDING OF RIBOSOMAL
CC       PROTEINS.
CC   -!- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC       PEPTIDE BONDS IN OLIGOPEPTIDES.
CC   -!- ENZYME REGULATION: INHIBITED BY BOTH FK506 AND RAPAMYCIN.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
CC   -!- PTM: PHOSPHORYLATED AT TYROSINE AND DEPHOSPHORYLATED BY THE
CC       PHOSPHOTYROSINE-SPECIFIC PHOSPHOPROTEIN PHOSPHATASE PTP1.
CC   -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L34569; AAB04165.1; -.
DR   EMBL; X79379; CAA55924.1; -.
DR   EMBL; S73876; AAB31995.1; -.
DR   EMBL; Z46373; CAA86504.1; -.
DR   PIR; S47927; S47927.
DR   PIR; A53544; A53544.
DR   HSSP; P20081; 1YAT.
DR   SGD; S0004539; NPI46.
DR   InterPro; IPR001179; FKBP_PPIase.
DR   Pfam; PF00254; FKBP; 1.
DR   PROSITE; PS00453; FKBP_PPIASE_1; 1.
DR   PROSITE; PS00454; FKBP_PPIASE_2; 1.
DR   PROSITE; PS50059; FKBP_PPIASE_3; 1.
KW   Isomerase; Rotamase; Nuclear protein; Phosphorylation.
FT   DOMAIN       60     87       ASP/GLU-RICH (HIGHLY ACIDIC).
FT   DOMAIN       88     99       LYS-RICH (HIGHLY BASIC).
FT   DOMAIN      101    119       ASP/GLU-RICH (HIGHLY ACIDIC).
FT   DOMAIN      173    248       ASP/GLU-RICH (HIGHLY ACIDIC).
FT   DOMAIN      250    298       LYS-RICH (HIGHLY BASIC).
FT   DOMAIN      256    271       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      324    411       PPIASE, FKBP-TYPE.
FT   MOD_RES     184    184       PHOSPHORYLATION (BY CK2).
FT   MOD_RES     186    186       PHOSPHORYLATION (BY CK2).
FT   CONFLICT    122    122       L -> P (IN REF. 1).
FT   CONFLICT    240    240       E -> EEE (IN REF. 1).
FT   CONFLICT    335    335       L -> F (IN REF. 1).
SQ   SEQUENCE   411 AA;  46553 MW;  A01D24DE0078FE11 CRC64;


  Query Match             12.5%;  Score 78;  DB 1;  Length 411;
  Best Local Similarity   24.8%;  Pred. No. 10;
  Matches   32;  Conservative   21;  Mismatches   60;  Indels   16;  Gaps    3;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           : :: ||:|  ||  ||::: ::|| : ||  :  | | :     ::        |: :|
Db     219 VQEEDEEDNDGEEEQEEEEEEEQKEEVKPEPKKSKKEKKRKHEEKEEE-------KKAKK 271

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHI----PTPQDLPQRKSSL-----V 111
            :|     |      | |:|:       |: |    |     |  | || |
Db     272 VKKVEFKKDLEEGPTKPKSKKEQDKHKPKSKVLEGGIVIEDRTIGDGPQAKRGARVGMRY 331
```

```
Qy     112 TSKLAGGQV 120
           ||  |:|
Db     332 IGKLKNGKV 340


RESULT 19
IF2P_HUMAN
ID   IF2P_HUMAN     STANDARD;      PRT;  1220 AA.
AC   O60841; O95805; Q9UF81; Q9UMN7;
DT   15-DEC-1998 (Rel. 37, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Translation initiation factor IF-2.
GN   IF2 OR KIAA0741.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99087487; PubMed=9872452;
RA   Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XI.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 5:277-286(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cervical carcinoma;
RX   MEDLINE=99362399; PubMed=10432305;
RA   Wilson S.A., Sieiro-Vazquez C., Edwards N.J., Iourin O., Byles E.D.,
RA   Kotsopoulou E., Adamson C.S., Kingsman S.M., Kingsman A.J.,
RA   Martin-Rendon E.;
RT   "Cloning and characterization of hIF2, a human homologue of bacterial
RT   translation initiation factor 2 and its interaction with HIV-1
RT   matrix.";
RL   Biochem. J. 342:97-103(1999).
RN   [3]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   TISSUE=Testis;
RX   MEDLINE=99218282; PubMed=10200264;
RA   Lee J.H., Choi S.K., Roll-Mecak A., Burley S.K., Dever T.E.;
RT   "Universal conservation in translation initiation revealed by human
RT   and archaeal homologs of bacterial translation initiation factor
RT   IF2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:4342-4347(1999).
RN   [4]
RP   SEQUENCE OF 89-1220 FROM N.A.
RC   TISSUE=Testis;
RA   Koehrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 833-1220 FROM N.A.
RX   PubMed=11124703;
RA   Stanchi F., Bertocco E., Toppo S., Dioguardi R., Simionati B.,
RA   Cannata N., Zimbello R., Lanfranchi G., Valle G.;
RT   "Characterization of 16 novel human genes showing high similarity to
RT   yeast sequences.";
RL   Yeast 18:69-80(2001).
CC   -!- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING
CC       THE BINDING OF THE FORMYLMETHIONINE-TRNA TO RIBOSOMES. SEEMS TO
CC       FUNCTION ALONG WITH EIF-2 (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB018284; BAA34461.1; -.
DR   EMBL; AJ006776; CAB44357.1; -.
DR   EMBL; AF078035; AAD16006.1; -.
DR   EMBL; AL133563; CAB63717.1; -.
DR   EMBL; AJ006412; CAA07018.1; -.
DR   MIM; 606086; -.
DR   InterPro; IPR000795; GTP_EFTU.
DR   InterPro; IPR004161; GTP_EFTU_D2.
DR   InterPro; IPR000178; IF2.
DR   Pfam; PF00009; GTP_EFTU; 1.
DR   Pfam; PF03144; GTP_EFTU_D2; 2.
DR   PRINTS; PR00315; ELONGATNFCT.
DR   PROSITE; PS01176; IF2; FALSE_NEG.
KW   Initiation factor; Protein biosynthesis; GTP-binding.
FT   NP_BIND     638    645       GTP (BY SIMILARITY).
FT   DOMAIN       39     50       POLY-LYS.
FT   DOMAIN       94     99       POLY-LYS.
FT   DOMAIN      138    142       POLY-ASP.
FT   DOMAIN      313    322       POLY-LYS.
FT   DOMAIN      353    356       POLY-GLU.
FT   DOMAIN      361    364       POLY-GLU.
FT   DOMAIN      491    496       POLY-GLU.
FT   DOMAIN      529    567       ASP/GLU-RICH (ACIDIC).
FT   MUTAGEN     640    640       V->G: LOSS OF ACTIVITY IN VIVO. RETAINS
FT                                FULL ACTIVITY IN VITRO.
FT   MUTAGEN     706    706       H->E: LOSS OF ACTIVITY; BOTH IN VIVO AND
FT                                IN VITRO.
FT   MUTAGEN     706    706       H->Q: LOSS OF ACTIVITY IN VIVO. PARTIAL
FT                                ACTIVITY IN VITRO.
FT   MUTAGEN     759    759       D->N: LOSS OF ACTIVITY; BOTH IN VIVO AND
FT                                IN VITRO.
FT   CONFLICT     64     64       E -> G (IN REF. 1).
FT   CONFLICT     92     92       T -> I (IN REF. 2).
FT   CONFLICT    180    180       I -> M (IN REF. 3).
FT   CONFLICT    256    256       K -> R (IN REF. 3).
FT   CONFLICT    522    522       T -> K (IN REF. 1).
FT   CONFLICT    549    549       E -> V (IN REF. 3).
FT   CONFLICT    669    669       G -> W (IN REF. 3).
FT   CONFLICT    894    894       E -> K (IN REF. 2).
SQ   SEQUENCE   1220 AA;  138799 MW;  010CC15DE475E5B8 CRC64;


  Query Match             12.3%;  Score 77;  DB 1;  Length 1220;
  Best Local Similarity   26.2%;  Pred. No. 40;
  Matches   28;  Conservative   21;  Mismatches   30;  Indels   28;  Gaps    6;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGIL--------PERAEEAKLKAKYPSLGQKPGGSDFL 54
           :::|||   ||  || :: :|:||           |:  : |   |  :| :||
Db     532 EEEEEEEEEEEEDEESEEEEEEEGESEGSEGDEEDEKVSDEKDSGK--TLDKKPS----- 584

Qy      55 MKRLQKGQKYFDSGD--------YNMAKAKMKNKQLPSAGPDKNLVT 93
             | :    :| || |        |: || ::: ::|  :   ||: |
Db     585 -KEMSSDSEY-DSDDDRTKEERAYDKAKRRIEKRRLEHS---KNVNT 626


RESULT 20
BASP_RAT
ID   BASP_RAT       STANDARD;      PRT;   219 AA.
AC   Q05175;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Brain acid soluble protein 1 (BASP1 protein) (Neuronal axonal membrane
DE   protein NAP-22).
GN   BASP1 OR NAP22.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
```

```
RC   TISSUE=Brain;
RX   MEDLINE=93293900; PubMed=8390468;
RA   Maekawa S., Maekawa M., Hattori S., Nakamura S.;
RT   "Purification and molecular cloning of a novel acidic calmodulin
RT   binding protein from rat brain.";
RL   J. Biol. Chem. 268:13703-13709(1993).
RN   [2]
RP   MYRISTOYLATION.
RX   MEDLINE=94250689; PubMed=8193160;
RA   Maekawa S., Matsuura Y., Nakaruma S.;
RT   "Expression and myristoylation of NAP-22 using a baculovirus transfer
RT   vector system.";
RL   Biochim. Biophys. Acta 1218:119-122(1994).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=97455984; PubMed=9310187;
RA   Mosevitsky M.I., Capony J.-P., Skladchikova G.Y.U., Novitskaya V.A.,
RA   Plekhanov A.Y.U., Zakharov V.V.;
RT   "The BASP1 family of myristoylated proteins abundant in axonal
RT   termini. Primary structure analysis and physico-chemical
RT   properties.";
RL   Biochimie 79:373-384(1997).
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ANCHORED. ASSOCIATED WITH THE
CC       MEMBRANES OF "GROWTH CONES" THAT FORM THE TIPS OF ELONGATING
CC       AXONS.
CC   -!- TISSUE SPECIFICITY: BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE BASP/NAP22 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D14441; BAA03333.1; -.
DR   PIR; A46597; A46597.
KW   Membrane; Myristate; Neurone; Calmodulin-binding.
FT   INIT_MET      0      0
FT   LIPID         1      1       MYRISTATE.
SQ   SEQUENCE   219 AA;  21659 MW;  7E79A441E021D163 CRC64;


  Query Match             12.2%;  Score 76.5;  DB 1;  Length 219;
  Best Local Similarity   25.0%;  Pred. No. 6.9;
  Matches   29;  Conservative   23;  Mismatches   47;  Indels   17;  Gaps    5;

Qy       5 QEEENPAEETGEEKQDTQ-EKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           ::::  ||  | |:: || | |      | | |  |:     :||       |    |:
Db      19 KDKDKKAEGAGTEEEGTQKESEPQAAADATEVKESAE-----EKP-------KDAADGEA 66

Qy      64 YFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAGGQ 119
               :  : : | ||   :: | | |:|:    :  | |   |:::::      |||:
Db      67 KAEEKEADKAAAK---EEAPKAEPEKSEGAAEEQPEPAPAPEQEAA-APGPAAGGE 118


RESULT  21
CALD_MELGA
ID   CALD_MELGA     STANDARD;      PRT;   239 AA.
AC   P13505;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Caldesmon, smooth muscle (CDM) (Fragment).
GN   CALD1 OR CAD.
OS   Meleagris gallopavo (Common turkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Meleagrididae; Meleagris.
OX   NCBI_TaxID=9103;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Gizzard;
RX   MEDLINE=89273603; PubMed=2730648;
RA   Leszyk J., Mornet D., Audemard E., Collins J.H.;
RT   "Caldesmon structure and function: sequence analysis of a 35
RT   kilodalton actin- and calmodulin-binding fragment from the C-terminus
RT   of the turkey gizzard protein.";
RL   Biochem. Biophys. Res. Commun. 160:1371-1378(1989).
RN   [2]
RP   SEQUENCE OF 1-96.
RC   TISSUE=Gizzard;
RX   MEDLINE=89228015; PubMed=2653315;
RA   Leszyk J., Mornet D., Audemard E., Collins J.H.;
RT   "Amino acid sequence of a 15 kilodalton actin-binding fragment of
RT   turkey gizzard caldesmon: similarity with dystrophin, tropomyosin and
RT   the tropomyosin-binding region of troponin T.";
RL   Biochem. Biophys. Res. Commun. 160:210-216(1989).
CC   -!- FUNCTION: CONTROL OF ACTOMYOSIN INTERACTIONS IN SMOOTH MUSCLE
CC       AND NONMUSCLE CELLS (COULD ACT AS A BRIDGE BETWEEN MYOSIN AND
CC       ACTIN FILAMENTS). INHIBITS THE ACTIN-ACTIVATED ATPASE OF MYOSIN
CC       THIS INHIBITION IS ATTENUATED BY CALCIUM-CALMODULIN AND IS
CC       POTENTIATED BY TROPOMYOSIN. INTERACTS WITH ACTIN, MYOSIN,
CC       2 MOLECULES OF TROPOMYOSIN AND WITH CALMODULIN.
CC   -!- SUBCELLULAR LOCATION: ON THIN FILAMENTS IN SMOOTH MUSCLE AND ON
CC       STRESS FIBERS IN FIBROBLASTS (NONMUSCLE).
DR   PIR; A32488; A32488.
DR   PIR; A32420; A32420.
KW   Muscle protein; Actin-binding; Calmodulin-binding.
FT   NON_TER       1      1
FT   NON_TER     239    239
SQ   SEQUENCE   239 AA;  26561 MW;  D395B8BB465B8895 CRC64;


  Query Match             12.2%;  Score 76.5;  DB 1;  Length 239;
  Best Local Similarity   29.2%;  Pred. No. 7.6;
  Matches   38;  Conservative   13;  Mismatches   40;  Indels   39;  Gaps    7;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLG----QKP-------GGS 51
             ||:::|    :  ||::  : || |   ||| |: : | |  |    :||       | |
Db      46 QKKKQEEAERKIREEEEKKRMKEEIERRRAEAAEKRQKMPEDGVSEEKKPFKCFSPKGSS 105

Qy      52 -------DFLMKRLQK-GQK---------YFDSGDYNMAKAKMKNKQLPSAGPDKNLVTG 94
                  :|| |  || | |           ||       | : ||   :| | |
Db     106 LKIEERAEFLNKSAQKSGMKPAHTTAVVSKIDSRLEQYTSAVVGNK---AAKPAK----- 157

Qy      95 DHIPTPQDLP 104
              |   |||
Db     158 ---PAASDLP 164


RESULT  22
CALD_CHICK
ID   CALD_CHICK     STANDARD;      PRT;   771 AA.
AC   P12957; Q90756; Q90761; Q92018; Q99230; Q03698;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Caldesmon (CDM).
GN   CALD1 OR CAD.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (GIZZARD H-CAD ISOFORM).
RC   TISSUE=Gizzard, and Oviduct;
RX   MEDLINE=89340480; PubMed=2760048;
RA   Bryan J., Imai M., Lee R., Moore P., Cook R.G., Lin W.-G.;
RT   "Cloning and expression of a smooth muscle caldesmon.";
RL   J. Biol. Chem. 264:13873-13879(1989).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (GIZZARD H-CAD ISOFORM).
RC   TISSUE=Gizzard;
```

```
RX   MEDLINE=90026426; PubMed=2803315;
RA   Hayashi K., Kanda K., Kimizuka F., Kato I., Sobue K.;
RT   "Primary structure and functional expression of h-caldesmon
RT   complementary DNA.";
RL   Biochem. Biophys. Res. Commun. 164:503-511(1989).
RN   [3]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (BRAIN L-CAD ISOFORM).
RC   TISSUE=Brain;
RX   MEDLINE=91093148; PubMed=1824698;
RA   Hayashi K., Fujio Y., Kato I., Sobue K.;
RT   "Structural and functional relationships between h- and
RT   l-caldesmons.";
RL   J. Biol. Chem. 266:355-361(1991).
RN   [4]
RP   SEQUENCE FROM N.A. (GIZZARD H-CAD; BRAIN L-CAD AND GIZZARD L-CAD).
RX   MEDLINE=94071934; PubMed=8250919;
RA   Haruna M., Hayashi K., Yano H., Takeuchi O., Sobue K.;
RT   "Common structural and expressional properties of vertebrate
RT   caldesmon genes.";
RL   Biochem. Biophys. Res. Commun. 197:145-153(1993).
RN   [5]
RP   SEQUENCE FROM N.A. (GIZZARD L-CAD ISOFORM).
RC   TISSUE=Gizzard;
RX   MEDLINE=92042686; PubMed=1939602;
RA   Bryan J., Lee R.;
RT   "Sequence of an avian non-muscle caldesmon.";
RL   J. Muscle Res. Cell Motil. 12:372-375(1991).
RN   [6]
RP   SEQUENCE OF 1-15 FROM N.A. (BRAIN L-CAD ISOFORM).
RC   TISSUE=Brain;
RX   MEDLINE=94271210; PubMed=8002994;
RA   Yano H., Hayashi K., Haruna M., Sobue K.;
RT   "Identification of two distinct promoters in the chicken caldesmon
RT   gene.";
RL   Biochem. Biophys. Res. Commun. 201:618-626(1994).
RN   [7]
RP   SEQUENCE OF 451-756 FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Gizzard;
RX   MEDLINE=89273666; PubMed=2730665;
RA   Hayashi K., Yamada S., Kanda K., Kimizuka F., Kato I., Sobue K.;
RT   "35 kDa fragment of h-caldesmon conserves two consensus sequences of
RT   the tropomyosin-binding domain in troponin T.";
RL   Biochem. Biophys. Res. Commun. 161:38-45(1989).
RN   [8]
RP   SEQUENCE OF 498-525.
RX   MEDLINE=88293484; PubMed=3401222;
RA   Mornet D., Audemard E., Derancourt J.;
RT   "Identification of a 15 kilodalton actin binding region on gizzard
RT   caldesmon probed by chemical cross-linking.";
RL   Biochem. Biophys. Res. Commun. 154:564-571(1988).
RN   [9]
RP   PHOSPHORYLATION SITES.
RX   MEDLINE=92041815; PubMed=1939059;
RA   Mak A.S., Carpenter M., Smillie L.B., Wang J.H.;
RT   "Phosphorylation of caldesmon by p34cdc2 kinase. Identification of
RT   phosphorylation sites.";
RL   J. Biol. Chem. 266:19971-19975(1991).
CC   -!- FUNCTION: ACTIN- AND MYOSIN-BINDING PROTEIN IMPLICATED IN THE
CC       REGULATION OF ACTOMYOSIN INTERACTIONS IN SMOOTH MUSCLE AND
CC       NONMUSCLE CELLS (COULD ACT AS A BRIDGE BETWEEN MYOSIN AND ACTIN
CC       FILAMENTS). STIMULATES ACTIN BINDING OF TROPOMYOSIN WHICH
CC       INCREASES THE STABILIZATION OF ACTIN FILAMENT STRUCTURE. IN MUSCLE
CC       TISSUES, INHIBITS THE ACTOMYOSIN ATPASE BY BINDING TO F-ACTIN.
CC       THIS INHIBITION IS ATTENUATED BY CALCIUM-CALMODULIN AND IS
CC       POTENTIATED BY TROPOMYOSIN. INTERACTS WITH ACTIN, MYOSIN, TWO
CC       MOLECULES OF TROPOMYOSIN AND WITH CALMODULIN. ALSO PLAY AN
CC       ESSENTIAL ROLE DURING CELLULAR MITOSIS AND RECEPTOR CAPPING.
CC   -!- SUBCELLULAR LOCATION: ON THIN FILAMENTS IN SMOOTH MUSCLE AND ON
CC       STRESS FIBERS IN FIBROBLASTS (NONMUSCLE) (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; GIZZARD H-CAD (SHOWN HERE),
CC       BRAIN L-CAD AND GIZZARD L-CAD; ARE PRODUCED BY ALTERNATIVE
CC       SPLICING.
CC   -!- TISSUE SPECIFICITY: HIGH-MOLECULAR-WEIGHT CALDESMON (H-CALDESMON)
CC       IS PREDOMINANTLY EXPRESSED IN SMOOTH MUSCLES, WHEREAS LOW-
CC       MOLECULAR-WEIGHT CALDESMON (L-CALDESMON) IS WIDELY DISTRIBUTED IN
CC       NON-MUSCLE TISSUES AND CELLS. NOT EXPRESSED IN SKELETAL MUSCLE OR
CC       HEART.
CC   -!- DOMAIN: THE N-TERMINAL PART SEEMS TO BE A MYOSIN/CALMODULIN-
CC       BINDING DOMAIN, AND THE C-TERMINAL A TROPOMYOSIN/ACTIN/CALMODULIN-
CC       BINDING DOMAIN. THESE TWO DOMAINS ARE SEPARATED BY A CENTRAL
CC       HELICAL REGION IN THE MUSCLE FORMS.
CC   -!- PTM: IN NON-MUSCLE CELLS, PHOSPHORYLATION BY CDC2 DURING MITOSIS
CC       CAUSES CALDESMON TO DISSOCIATE FROM MICROFILAMENTS.
CC       PHOSPHORYLATION REDUCES CALDESMON BINDING TO ACTIN, MYOSIN, AND
CC       CALMODULIN AS WELL AS ITS INHIBITION OF ACTOMYOSIN ATPASE
CC       ACTIVITY. PHOSPHORYLATION ALSO OCCURS IN BOTH QUIESCENT AND
CC       DIVIDING SMOOTH MUSCLE CELLS WITH SIMILAR EFFECTS ON THE
CC       INTERACTION WITH ACTIN AND CALMODULIN AND ON MICROFILAMENTS
CC       REORGANIZATION (BY SIMILARITY).
CC   -!- SIMILARITY: TO A TROPOMYOSIN BINDING SITE DOMAIN OF TROPONIN T.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04968; AAA49067.1; -.
DR   EMBL; D17648; BAA04539.1; -.
DR   EMBL; D17634; BAA04539.1; JOINED.
DR   EMBL; D17636; BAA04539.1; JOINED.
DR   EMBL; D17637; BAA04539.1; JOINED.
DR   EMBL; D17638; BAA04539.1; JOINED.
DR   EMBL; D17639; BAA04539.1; JOINED.
DR   EMBL; D17640; BAA04539.1; JOINED.
DR   EMBL; D17641; BAA04539.1; JOINED.
DR   EMBL; D17642; BAA04539.1; JOINED.
DR   EMBL; D17643; BAA04539.1; JOINED.
DR   EMBL; D17644; BAA04539.1; JOINED.
DR   EMBL; D17645; BAA04539.1; JOINED.
DR   EMBL; D17646; BAA04539.1; JOINED.
DR   EMBL; M28417; AAA48810.1; -.
DR   EMBL; M60620; AAA48936.1; -.
DR   EMBL; D17648; BAA04538.1; -.
DR   EMBL; D17635; BAA04538.1; JOINED.
DR   EMBL; D17636; BAA04538.1; JOINED.
DR   EMBL; D17637; BAA04538.1; JOINED.
DR   EMBL; D17639; BAA04538.1; JOINED.
DR   EMBL; D17640; BAA04538.1; JOINED.
DR   EMBL; D17641; BAA04538.1; JOINED.
DR   EMBL; D17642; BAA04538.1; JOINED.
DR   EMBL; D17643; BAA04538.1; JOINED.
DR   EMBL; D17644; BAA04538.1; JOINED.
DR   EMBL; D17645; BAA04538.1; JOINED.
DR   EMBL; D17646; BAA04538.1; JOINED.
DR   EMBL; D17647; BAA04538.1; JOINED.
DR   EMBL; D17648; BAA04540.1; -.
DR   EMBL; D17634; BAA04540.1; JOINED.
DR   EMBL; D17636; BAA04540.1; JOINED.
DR   EMBL; D17637; BAA04540.1; JOINED.
DR   EMBL; D17639; BAA04540.1; JOINED.
DR   EMBL; D17640; BAA04540.1; JOINED.
DR   EMBL; D17641; BAA04540.1; JOINED.
DR   EMBL; D17642; BAA04540.1; JOINED.
DR   EMBL; D17643; BAA04540.1; JOINED.
DR   EMBL; D17644; BAA04540.1; JOINED.
DR   EMBL; D17645; BAA04540.1; JOINED.
DR   EMBL; D17646; BAA04540.1; JOINED.
DR   EMBL; D17647; BAA04540.1; JOINED.
DR   EMBL; M59762; AAA48649.1; -.
DR   EMBL; D17552; BAA04490.1; -.
DR   EMBL; M26684; AAA48811.1; -.
DR   PIR; A32642; A32642.
DR   HSSP; P05412; 1FOS.
```

```
DR   InterPro; IPR000075; Caldesmon.
DR   Pfam; PF02029; Caldesmon; 1.
DR   PRINTS; PR01076; CALDESMON.
KW   Muscle protein; Actin-binding; Calmodulin-binding; Phosphorylation;
KW   Repeat; Alternative splicing.
FT   DOMAIN       26    199       MYOSIN AND CALMODULIN-BINDING.
FT   DOMAIN      251    390       10 X 13 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT      251    265       1.
FT   REPEAT      266    278       2.
FT   REPEAT      279    291       3.
FT   REPEAT      294    306       4.
FT   REPEAT      309    321       5.
FT   REPEAT      324    336       6.
FT   REPEAT      337    349       7.
FT   REPEAT      350    362       8.
FT   REPEAT      363    375       9.
FT   REPEAT      378    390       10.
FT   DOMAIN      523    580       TROPOMYOSIN-BINDING (POTENTIAL).
FT   DOMAIN      622    632       TROPOMYOSIN-BINDING (POTENTIAL).
FT   DOMAIN      612    644       STRONG ACTIN-BINDING.
FT   DOMAIN      674    680       CALMODULIN-BINDING.
FT   DOMAIN      726    752       WEAK ACTIN-BINDING.
FT   DOMAIN       39     46       POLY-ARG.
FT   DOMAIN      539    542       POLY-GLU.
FT   DOMAIN      556    559       POLY-GLU.
FT   MOD_RES     485    485       PHOSPHORYLATION (POTENTIAL).
FT   MOD_RES     597    597       PHOSPHORYLATION (BY CDC2).
FT   MOD_RES     682    682       PHOSPHORYLATION (BY CDC2).
FT   MOD_RES     688    688       PHOSPHORYLATION (BY CDC2).
FT   MOD_RES     711    711       PHOSPHORYLATION (BY CDC2).
FT   MOD_RES     717    717       PHOSPHORYLATION (BY CDC2).
FT   VARSPLIC      1     24       MDDFERRRELRRQKREEMRLEAER -> MISRSYCRQNLS

  Query Match             12.2%;  Score 76.5;  DB 1;  Length 771;
  Best Local Similarity   29.2%;  Pred. No. 27;
  Matches   38;  Conservative   13;  Mismatches   40;  Indels   39;  Gaps    7;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLG----QKP-------GGS 51
           ||:::|    :  ||::  : || |   ||| |: : | |  |    :||       | |
Db     543 QKKKQEEAERKIREEEEKKRMKEEIERRRAEAAEKRQKVPEDGVSEEKKPFKCFSPKGSS 602

Qy      52 -------DFLMKRLQK-GQK---------YFDSGDYNMAKAKMKNKQLPSAGPDKNLVTG 94
                :|| |  || | |           ||       | : ||   :| | |
Db     603 LKIEERAEFLNKSAQKSGMKPAHTTAVVSKIDSRLEQYTSAVVGNK---AAKPAK----- 654

Qy      95 DHIPTPQDLP 104
             |   |||
Db     655 ---PAASDLP 661


RESULT  23
BRD2_HUMAN
ID   BRD2_HUMAN     STANDARD;      PRT;   801 AA.
AC   P25440;
DT   01-MAY-1992 (Rel. 22, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Bromodomain-containing protein 2 (RING3 protein).
GN   BRD2 OR RING3 OR KIAA9001.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=T-cell;
RX   MEDLINE=92329974; PubMed=1352711;
RA   Beck S., Hanson I., Kelly A., Pappin D.J.C., Trowsdale J.;
RT   "A homologue of the Drosophila female sterile homeotic (fsh) gene in
RT   the class II region of the human MHC.";
RL   DNA Seq. 2:203-210(1992).
RN   [2]
RP   REVISIONS TO N-TERMINUS.
RX   MEDLINE=96376536; PubMed=8781126;
RA   Thorpe K.L., Abdulla S., Kaufman J., Trowsdale J., Beck S.;
RT   "Phylogeny and structure of the RING3 gene.";
RL   Immunogenetics 44:391-396(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RA   Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayasi Y.,
RA   Sato S., Nagase T., Seki T., Ishikawa K.-I., Tabata S.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 ET DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X62083; CAA43996.1; -.
DR   EMBL; M80613; AAA68890.1; ALT_INIT.
DR   EMBL; D42040; BAA07641.1; -.
DR   PIR; S18860; S18860.
DR   PIR; S40781; S40781.
DR   MIM; 601540; -.
DR   InterPro; IPR001487; Bromodomain.
DR   Pfam; PF00439; bromodomain; 2.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   SMART; SM00297; BROMO; 2.
DR   PROSITE; PS00633; BROMODOMAIN_1; 2.
DR   PROSITE; PS50014; BROMODOMAIN_2; 2.
KW   Bromodomain; Repeat; Nuclear protein.
FT   DOMAIN       91    163       BROMODOMAIN 1.
FT   DOMAIN      364    436       BROMODOMAIN 2.
FT   DOMAIN      476    515       GLU/SER-RICH.
FT   DOMAIN      544    566       ARG/LYS-RICH (HIGHLY BASIC).
FT   DOMAIN      775    801       SER-RICH.
FT   DOMAIN      555    559       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      638    801       ET DOMAIN.
FT   DOMAIN       61     64       POLY-PRO.
FT   DOMAIN      492    506       POLY-GLU.
FT   DOMAIN      551    559       POLY-LYS.
FT   DOMAIN      634    638       POLY-GLU.
FT   DOMAIN      775    793       POLY-SER.
SQ   SEQUENCE   801 AA;  88060 MW;  9A075EEB13507D8E CRC64;


  Query Match             12.2%;  Score 76.5;  DB 1;  Length 801;
  Best Local Similarity   28.2%;  Pred. No. 28;
  Matches   35;  Conservative   17;  Mismatches   47;  Indels   25;  Gaps    6;

Qy       2 SQKQEEENPAEETGEEKQDTQEKEGILPERAE-----EAKLKAKYPSLGQKPGG--SDFL 54
           |:::|||:  :|  || : :  :|    |||      : :|:| :  |     |  |
Db     491 SEEEEEEDEEDEEEEESESSDSEE----ERAHRLAELQEQLRAVHEQLAALSQGPISKPK 546

Qy      55 MKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSK 114
            || :| :|          | :   |    || |::   |   | |   || | |   |
Db     547 RKREKKEKK----------KKRKAEKHRGRAGADED-DKGPRAPRP---PQPKKSKKASG 592

Qy     115 LAGG 118
             ||
Db     593 SGGG 596


RESULT  24
HAP1_RAT
ID   HAP1_RAT       STANDARD;      PRT;   629 AA.
AC   P54256;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
```

```
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Huntingtin associated protein 1 (HAP1).
GN   HAP1-A AND HAP1-B.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=96077112; PubMed=7477378;
RA   Li X.-J., Li S.-H., Sharp A.H., Nucifora F.C. Jr., Schilling G.,
RA   Lanahan A., Worley P., Snyder S.H., Ross C.A.;
RT   "A huntingtin-associated protein enriched in brain with implications
RT   for pathology.";
RL   Nature 378:398-402(1995).
CC   -!- FUNCTION: ASSOCIATES SPECIFICALLY WITH HUNTINGTIN. THIS BINDING IS
CC       ENHANCED BY AN EXPANDED POLYGLUTAMINE REPEAT.
CC   -!- ALTERNATIVE PRODUCTS: ALTERNATIVE SPLICING PRODUCES A SHORT FORM
CC       (A) AND A LONG FORM (B). THE LATEST IS SHOWN HERE.
CC   -!- TISSUE SPECIFICITY: IN THE BRAIN, SPECIALLY IN THE OLFACTORY BULB
CC       AND IN THE BRAIN STEM. NO DETECTABLE EXPRESSION IN PERIPHERAL
CC       TISSUES SUCH AS LUNG, TESTIS, SPLEEN, AND SMALL INTESTINE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U38370; AAC52326.1; -.
DR   EMBL; U38373; AAC52327.1; -.
KW   Alternative splicing.
FT   DOMAIN      220    237       POLY-GLU.
FT   DOMAIN      468    471       POLY-PRO.
FT   VARSPLIC    579    599       DSPAPQQQTNMGGGIVEQQPI -> GECSRRGHPPASGTSY
FT                                RSSTL (IN ISOFORM A).
FT   VARSPLIC    600    629       MISSING (IN ISOFORM A).
SQ   SEQUENCE   629 AA;  70213 MW;  C23C5CB57520B574 CRC64;


  Query Match             12.2%;  Score 76;  DB 1;  Length 629;
  Best Local Similarity   23.3%;  Pred. No. 24;
  Matches   28;  Conservative   22;  Mismatches   38;  Indels   32;  Gaps    5;

Qy       7 EENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL---GQKPGGSD--FLMKRLQKG 61
           |: ||||| :  ::|:  | : ||  |  ::     :|   |  |   |  :::::|
Db     503 EDGPAEETEQASEETEAWEEVEPEVDEATRMNVVVSALEASGLGPSHLDMKYVLQQLSNW 562

Qy      62 QKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAGGQVE 121
           |          :| : | | :|   |                ||::::     : || ||
Db     563 QDAH-------SKRQQKQKVVPKDSP---------------APQQQTN-----MGGGIVE 595


RESULT  25
TRT2_RABIT
ID   TRT2_RABIT     STANDARD;      PRT;   300 AA.
AC   P09741; P09742;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Troponin T, cardiac muscle isoforms (TnTC).
GN   TNNT2.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=94084880; PubMed=8261593;
RA   Greig A., Hirschberg Y., Anderson P.A.W., Hainsworth C.,
RA   Malouf N.N., Oakeley A.E., Kay B.K.;
RT   "Molecular basis of cardiac troponin T isoform heterogeneity in
RT   rabbit heart.";
RL   Circ. Res. 74:41-47(1994).
RN   [2]
RP   SEQUENCE OF 1-15; 22; 33-206 AND 216-300.
RX   MEDLINE=87057385; PubMed=3782144;
RA   Pearlstone J.R., Carpenter M.R., Smillie L.B.;
RT   "Amino acid sequence of rabbit cardiac troponin T.";
RL   J. Biol. Chem. 261:16795-16810(1986).
CC   -!- FUNCTION: TROPONIN T IS THE TROPOMYOSIN-BINDING SUBUNIT OF
CC       TROPONIN, THE THIN FILAMENT REGULATORY COMPLEX WHICH CONFERS
CC       CALCIUM-SENSITIVITY TO STRIATED MUSCLE ACTOMYOSIN ATPASE ACTIVITY.
CC   -!- SUBUNIT: BINDS WITH TROPONINS I AND C TO MAKE THE THIN-FILAMENT
CC       REGULATORY COMPLEX, TROPONIN.
CC   -!- ALTERNATIVE PRODUCTS: THE DIFFERENT ISOFORMS ARE GENERATED BY A
CC       DIFFERENTIAL MRNA SPLICING MECHANISM. MYOFILAMENTS CONTAINING A
CC       RELATIVELY GREATER AMOUNT OF A HIGH MOLECULAR WEIGHT CTNT ISOFORM
CC       BOUND LESS CALCIUM TO MYOFIBRILLAR TNC THAN MYOFILAMENTS THAT DID
CC       NOT CONTAIN THE HIGH MOLECULAR WEIGHT CTNT. THE MAJOR ISOFORM IN
CC       ADULT HEART IS CTNT4.
CC   -!- MISCELLANEOUS: THE SEQUENCE OF ISOFORM CTNT1 IS SHOWN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L40178; AAB51160.1; ALT_INIT.
DR   PIR; A25345; A25345.
DR   PIR; B25345; B25345.
DR   InterPro; IPR001978; Troponin.
DR   Pfam; PF00992; Troponin; 1.
KW   Muscle protein; Alternative splicing; Multigene family;
KW   Phosphorylation.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       PHOSPHORYLATION (BY CK2)
FT                                (BY SIMILARITY).
FT   VARSPLIC     16     21       MISSING (IN ISOFORM CTNT2 AND ISOFORM
FT                                CTNT4).
FT   VARSPLIC     23     32       MISSING (IN ISOFORM CTNT3 AND ISOFORM
FT                                CTNT4).
FT   CONFLICT     45     45       R -> A (IN REF. 2).
FT   CONFLICT    138    138       D -> E (IN REF. 2).
FT   CONFLICT    146    146       D -> ER (IN REF. 2).
FT   CONFLICT    183    183       E -> D (IN REF. 2).
FT   CONFLICT    194    194       MISSING (IN REF. 1).
SQ   SEQUENCE   300 AA;  35745 MW;  DA97C09EA37169A5 CRC64;


  Query Match             12.1%;  Score 75.5;  DB 1;  Length 300;
  Best Local Similarity   29.4%;  Pred. No. 12;
  Matches   20;  Conservative   17;  Mismatches   18;  Indels   13;  Gaps    3;

Qy       2 SQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAK--YPSL-------GQKPGGSD 52
           :: : ||  ||| |:|::| ::::|      ||:| | :   |:|       |::    |
Db      47 AEAETEETQAEEDGQEEEDKEDEDG----PVEESKPKPRPFMPNLVPPKIPDGERVDFDD 102

Qy      53 FLMKRLQK 60
             ||::|
Db     103 IHRKRMEK 110


RESULT  26
SMY2_YEAST
ID   SMY2_YEAST     STANDARD;      PRT;   790 AA.
AC   P32909;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
```

```
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   SMY2 protein.
GN   SMY2 OR YBR172C OR YBR1233.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lillie S.H., Brown S.S.;
RL   Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 323-790 FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=94025982; PubMed=8212898;
RA   Schaaff-Gerstenschlaeger I., Bauer A., Boles E., Zimmermann F.K.;
RT   "Sequence and function analysis of a 4.3 kb fragment of Saccharomyces
RT   cerevisiae chromosome II including three open reading frames.";
RL   Yeast 9:915-921(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RA   Entian K.-D., Koetter P., Rose M., Becker J., Grey M., Li Z.,
RA   Niegemann E., Schenk-Groeninger R., Servos J., Wehner E.,
RA   Wolter R., Brendel M., Bauer J., Braun H., Dern K., Duesterhus S.,
RA   Gruenbein R., Hedges D., Kiesau P., Korol S., Krems B., Proft M.,
RA   Siegers K., Baur A., Boles E., Miosga T.,
RA   Schaaff-Gerstenschlaeger I., Zimmermann F.K.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: SUPPRESSOR OF THE MYO2 GENE.
CC   -!- SIMILARITY: SOME, TO S.POMBE SPAC4F10.13C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M90654; AAA35057.1; -.
DR   EMBL; X74437; CAA52452.1; -.
DR   EMBL; Z36041; CAA85133.1; -.
DR   PIR; S27458; S27458.
DR   SGD; S0000376; SMY2.
DR   InterPro; IPR003169; GYF.
DR   Pfam; PF02213; GYF; 1.
DR   SMART; SM00444; GYF; 1.
SQ   SEQUENCE   790 AA;  87098 MW;  99C0547FD1E68366 CRC64;


  Query Match             12.1%;  Score 75.5;  DB 1;  Length 790;
  Best Local Similarity   29.4%;  Pred. No. 34;
  Matches   35;  Conservative   13;  Mismatches   48;  Indels   23;  Gaps    5;

Qy       2 SQKQE----------EENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGS 51
           | |||          ::  :|   :   | |||     :| |||:|  |     ||
Db     402 SHKQEYKSTTMLEKGKKEKSESVAKALLDEQEKRNRELKRKEEARLSKK-----QKQKED 456

Qy      52 DFLMK-RLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSS 109
           | | | : || ||  :: :    |   | |:      |    | :   | :|||   ||
Db     457 DLLKKQKEQKEQKEKEALEAEKQKKSEKTKK------DTQTQT-EGFKTSKDLPSLNSS 508


RESULT  27
NFM_RAT
ID   NFM_RAT        STANDARD;      PRT;   845 AA.
AC   P12839; Q63370;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neurofilament triplet M protein (160 kDa neurofilament protein)
DE   (Neurofilament medium polypeptide) (NF-M).
GN   NEF3 OR NEFM OR NFM.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87282618; PubMed=2441012;
RA   Napolitano E.W., Chin S.S.M., Colman D.R., Liem R.K.H.;
RT   "Complete amino acid sequence and in vitro expression of rat NF-M,
RT   the middle molecular weight neurofilament protein.";
RL   J. Neurosci. 7:2590-2599(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR;
RX   MEDLINE=92332596; PubMed=1321159;
RA   Kelly B.M., Gillespie C.S., Sherman D.L., Brophy P.J.;
RT   "Schwann cells of the myelin-forming phenotype express neurofilament
RT   protein NF-M.";
RL   J. Cell Biol. 118:397-410(1992).
RN   [3]
RP   PHOSPHORYLATION SITES, AND REVISION TO 500.
RX   MEDLINE=92165797; PubMed=1537832;
RA   Xu Z.-S., Liu W.-S., Willard M.B.;
RT   "Identification of six phosphorylation sites in the COOH-terminal
RT   tail region of the rat neurofilament protein M.";
RL   J. Biol. Chem. 267:4467-4471(1992).
RN   [4]
RP   CARBOHYDRATE-LINKAGE SITES.
RX   MEDLINE=93346421; PubMed=8344946;
RA   Dong D.L.-Y., Xu Z.-S., Chevrier M.R., Cotter R.J., Cleveland D.W.,
RA   Hart G.W.;
RT   "Glycosylation of mammalian neurofilaments. Localization of multiple
RT   O-linked N-acetylglucosamine moieties on neurofilament polypeptides
RT   L and M.";
RL   J. Biol. Chem. 268:16679-16687(1993).
CC   -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC       AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC   -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
CC       PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC       THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC       INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC       OF AXONAL CALIBER.
CC   -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC       OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC       LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC       COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M18628; AAA41696.1; -.
DR   EMBL; Z12152; CAA78136.1; -.
DR   GlycoSuiteDB; P12839; -.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW   Glycoprotein.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1    103       HEAD.
FT   DOMAIN      103    410       ROD.
FT   DOMAIN      411    844       TAIL.
FT   DOMAIN      103    134       COIL 1A.
FT   DOMAIN      135    147       LINKER 1.
FT   DOMAIN      148    246       COIL 1B.
FT   DOMAIN      247    263       LINKER 12.
```

```
FT   DOMAIN       264    285       COIL 2A.
FT   DOMAIN       286    289       LINKER 2.
FT   DOMAIN       290    410       COIL 2B.
FT   CARBOHYD      47     47       O-LINKED (GLCNAC).
FT                                  /FTId=CAR_000130.
FT   CARBOHYD     430    430       O-LINKED (GLCNAC).
FT                                  /FTId=CAR_000131.
FT   MOD_RES      502    502       PHOSPHORYLATION.
FT   MOD_RES      506    506       PHOSPHORYLATION.
FT   MOD_RES      536    536       PHOSPHORYLATION.
FT   MOD_RES      603    603       PHOSPHORYLATION.
FT   MOD_RES      608    608       PHOSPHORYLATION.
FT   MOD_RES      666    666       PHOSPHORYLATION.
FT   CONFLICT      17     17       MISSING (IN REF. 2).
FT   CONFLICT      21     21       R -> P (IN REF. 2).
FT   CONFLICT     204    204       V -> L (IN REF. 2).
FT   CONFLICT     500    500       MISSING (IN REF. 1).
SQ   SEQUENCE   845 AA;  95660 MW;  316C41655B11197D CRC64;


  Query Match             12.1%;  Score 75.5;  DB 1;  Length 845;
  Best Local Similarity   22.9%;  Pred. No. 36;
  Matches   25;  Conservative   23;  Mismatches   58;  Indels    3;  Gaps    2;

Qy       3 QKQEEENPAEETGEEKQDTQEK--EGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           |::| |  ||  ||| :  :||  || : | | : ::| : |   : |     : :   |
Db     557 QEEEGETEAEGEGEEAEAKEEKKTEGKVEEMAIKEEIKVEKPEKAKSPVPKSPVEEVKPK 616

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSS 109
             :    :   :  : |::   | :  |      |:|:| :| :
Db     617 PEAKAGKDEQKEEEKVEEKKEVAKESPKEEKVEKKE-EKPKDVPDKKKA 664


RESULT  28
NRT_DROME
ID   NRT_DROME      STANDARD;      PRT;   846 AA.
AC   P23654;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Neurotactin.
GN   NRT.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91301057; PubMed=2100266;
RA   Hortsch M., Patel N.P., Bieber A.J., Traquina Z.R., Goodman C.S.;
RT   "Drosophila neurotactin, a surface glycoprotein with homology to
RT   serine esterases, is dynamically expressed during embryogenesis.";
RL   Development 110:1327-1340(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R; TISSUE=Embryo;
RX   MEDLINE=91006059; PubMed=2120047;
RA   de la Escalera S., Bockamp E.O., Moya F., Piovant M., Jimenez F.;
RT   "Characterization and gene cloning of neurotactin, a Drosophila
RT   transmembrane protein related to cholinesterases.";
RL   EMBO J. 9:3593-3601(1990).
RN   [3]
RP   FUNCTION.
RX   MEDLINE=91006060; PubMed=2120048;
RA   Barthalay Y., Hipeau-Jacquotte R., de la Escalera S., Jimenez F.,
RA   Piovant M.;
RT   "Drosophila neurotactin mediates heterophilic cell adhesion.";
RL   EMBO J. 9:3603-3609(1990).
CC   -!- FUNCTION: MAY MEDIATE OR MODULATE CELL ADHESION BETWEEN EMBRYONIC
CC       CELLS DURING DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN MEMBRANES DURING EMBRYOGENESIS
CC       AND THEN ACCUMULATES IN CNS AND PNS.
CC   -!- PTM: PHOSPHORYLATED.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE TYPE-B
CC       CARBOXYLESTERASE/LIPASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53837; CAA37831.1; -.
DR   EMBL; X54999; CAA38746.1; -.
DR   PIR; S12005; S12005.
DR   PIR; S13795; S13795.
DR   HSSP; P37967; 1QE3.
DR   FlyBase; FBgn0004108; Nrt.
DR   InterPro; IPR002018; Carboxylesterase_B.
DR   InterPro; IPR000379; Est_lip_thioest_actsite.
DR   Pfam; PF00135; COesterase; 1.
DR   PROSITE; PS00941; CARBOXYLESTERASE_B_2; 1.
KW   Cell adhesion; Glycoprotein; Transmembrane; Phosphorylation;
KW   Signal-anchor.
FT   DOMAIN         1    324       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     325    346       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                  (POTENTIAL).
FT   DOMAIN       347    846       EXTRACELLULAR (POTENTIAL).
FT   SIMILAR      351    846       WITH TYPE-B CARBOXYLESTERASE/LIPASES.
FT   MOD_RES       28     28       PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT   MOD_RES       75     75       PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT   MOD_RES      103    103       PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT   MOD_RES      169    169       PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT   MOD_RES      186    186       PHOSPHORYLATION (POTENTIAL).
FT   DISULFID     422    437       BY SIMILARITY.
FT   DISULFID     600    605       BY SIMILARITY.
FT   DISULFID     738    830       BY SIMILARITY.
FT   CARBOHYD     410    410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     428    428       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     636    636       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     691    691       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     720    720       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     250    250       G -> S (IN REF. 2).
FT   CONFLICT     555    555       S -> T (IN REF. 2).
FT   CONFLICT     567    567       F -> Y (IN REF. 2).
SQ   SEQUENCE   846 AA;  92745 MW;  F5EC54904D1D0657 CRC64;


  Query Match             12.1%;  Score 75.5;  DB 1;  Length 846;
  Best Local Similarity   24.4%;  Pred. No. 36;
  Matches   29;  Conservative   21;  Mismatches   50;  Indels   19;  Gaps    5;

Qy       1 MSQKQEEENPAEETGEEKQDT----QEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMK 56
           | : :|:| |  ||   :|:     :| : :| :: |:|| |   |    ||   :   |
Db       1 MGELEEKETPPTETTAAQQEALEEPKETDKML-DKKEDAKEKTPSPQT-SKPASPNAGKK 58

Qy      57 RLQKGQKYFDSGDYNMAKAK----------MKNKQLPSAGPDKNLVTGDHIPTPQDLP 104
             :|  |  |  :||:|            :|   | :  || :   :    |: :|
Db      59 SSPVAEKKID--DAELAKSKSGNGEEIIDIPAENGTKPDSADDKKISKEEREVKPKKIP 115


RESULT  29
NUKS_RAT
ID   NUKS_RAT       STANDARD;      PRT;   243 AA.
AC   Q9EPJ0;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nuclear ubiquitous casein and cyclin-dependent kinases substrate.
GN   NUCKS.
OS   Rattus norvegicus (Rat).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RA   Oestvold A.C., Wanvik B., Mathiesen S., Norum J.H., Nordeide H.,
RA   Olstad O.K.;
RT   "Molecular cloning of a novel mammalian nuclear phosphoprotein serving
RT   as substrate for CDK1 in vivo.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- PTM: PHOSPHORYLATED BY CDK1 AND CASEIN KINASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ237669; CAC20862.1; -.
KW   Phosphorylation; Nuclear protein.
FT   DOMAIN       140    145       POLY-ASP.
FT   DOMAIN       150    206       LYS-RICH.
SQ   SEQUENCE   243 AA;  27140 MW;  98B5BB2A7F2133C8 CRC64;


  Query Match             12.0%;  Score 75;  DB 1;  Length 243;
  Best Local Similarity   23.1%;  Pred. No. 10;
  Matches   33;  Conservative   27;  Mismatches   35;  Indels   48;  Gaps    8;

Qy       4 KQEEENPAEETGEEKQD--------------TQEKEGIL--------PERAEEAKLKAK 40
           |::: : ||:: :|| |               ::::| :|        ||  :||  :
Db      69 KKDDSHSAEDSEDEKDDHKSVRQQRQAASKAASKQREMLLEDVGSEEEPEEDDEAPFQ-- 126

Qy      41 YPSLGQKPGGS--DFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIP 98
                :|  ||  ||||:         |  ||  :| | | | :  : |::
Db     127 -----EKDSGSDEDFLMED-------DDDSDYGSSKKKNK-KMVKKSKPERK-------E 166

Qy      99 TPQDLPQRKSSLVTSKLAG-GQV 120
               |: |:::  | : | |:|
Db     167 KKMPKPRLKATVTPSPVKGKGKV 189


RESULT  30
FKB4_DROME
ID   FKB4_DROME     STANDARD;      PRT;   357 AA.
AC   P54397; Q9VF88;
DT   01-OCT-1996 (Rel. 34, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   39 kDa FK506-binding nuclear protein (Peptidyl-prolyl cis-trans
DE   isomerase) (PPiase) (EC 5.2.1.8).
GN   FK506-BP1 OR FKBP39 or CG6226.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE=95278752; PubMed=7538962;
RA   Theopold U., Dal Zotto L., Hultmark D.;
RT   "FKBP39, a Drosophila member of a family of proteins that bind the
RT   immunosuppressive drug FK506.";
RL   Gene 156:247-251(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS. FKBP506
CC       MAY FUNCTION IN A SIGNAL TRANSDUCTION CASCADE DURING EARLY
CC       DEVELOPMENT.
CC   -!- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC       PEPTIDE BONDS IN OLIGOPEPTIDES.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED, HIGHEST LEVELS IN
CC       OVARY.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED DURING ALL STAGES OF DEVELOPMENT
CC       WITH HIGHEST EXPRESSION IN EARLY EMBRYO.
CC   -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z46894; CAA86996.1; -.
DR   EMBL; AE003708; AAF55172.1; ALT_SEQ.
DR   HSSP; P27124; 1ROT.
DR   FlyBase; FBgn0013269; FK506-bp1.
DR   InterPro; IPR001179; FKBP_PPIase.
DR   Pfam; PF00254; FKBP; 1.
DR   PROSITE; PS00453; FKBP_PPIASE_1; FALSE_NEG.
DR   PROSITE; PS00454; FKBP_PPIASE_2; 1.
DR   PROSITE; PS50059; FKBP_PPIASE_3; 1.
KW   Isomerase; Rotamase; Nuclear protein.
FT   DOMAIN       89     99       ASP/GLU-RICH (HIGHLY ACIDIC).
FT   DOMAIN      119    183       ASP/GLU-RICH (HIGHLY ACIDIC).
FT   DOMAIN      186    247       LYS-RICH (BASIC).
FT   DOMAIN      269    357       PPIASE, FKBP-TYPE.
```

```
FT   CONFLICT    187    187       A -> R (IN REF. 1).
SQ   SEQUENCE   357 AA;  39343 MW;  EF0AB7831738BB30 CRC64;


  Query Match             12.0%;  Score 75;  DB 1;  Length 357;
  Best Local Similarity   32.4%;  Pred. No. 16;
  Matches   33;  Conservative   10;  Mismatches   35;  Indels   24;  Gaps    6;

Qy       6 EEENPAEETGEE-KQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKG--- 61
           :|:|  ||:||| :||: : |      ||| : |||   |   || |     : | |
Db     161 DEDN--EESGEEDEQDSDDSE------AEEEQPKAKVAKL--SPGASAKKSGKEQNGVAK 210

Qy      62 ---------QKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTG 94
                    :|          |||   || | |  |   :||
Db     211 KEEAKQQQKKKEKPEAKKEQPKAKEPAKQQP-ASKDPRTITG 251


RESULT  31
HPR1_YEAST
ID   HPR1_YEAST     STANDARD;      PRT;   752 AA.
AC   P17629; Q03918;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HPR1 protein.
GN   HPR1 OR YDR138W OR YD9302.14.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90205821; PubMed=2181275;
RA   Aguilera A., Klein H.L.;
RT   "HPR1, a novel yeast gene that prevents intrachromosomal excision
RT   recombination, shows carboxy-terminal homology to the Saccharomyces
RT   cerevisiae TOP1 gene.";
RL   Mol. Cell. Biol. 10:1439-1451(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Oliver K., Harris D., Barrell B.G., Rajandream M.A.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HPR1 MAY HAVE A TOP1-RELATED FUNCTION THAT ACTS TO
CC       PREVENT INTRACHROMOSOMAL EXCISION RECOMBINATION BETWEEN REPEATS,
CC       RECOMBINATION THAT LEADS TO DELETIONS OF INTERSTITIAL SEQUENCES.
CC   -!- MISCELLANEOUS: HPR1 IS NOT REQUIRED FOR SPORULATION, AND IS NOT
CC       ESSENTIAL FOR DNA REPAIR. IT IS PROBABLY NOT A DNA TOPOISOMERASE.
CC   -!- SIMILARITY: SHOWS C-TERMINAL HOMOLOGY TO YEAST TOP1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M30484; AAA34685.1; -.
DR   EMBL; Z48179; CAA88220.1; -.
DR   PIR; A34781; A34781.
DR   SGD; S0002545; HPR1.
FT   SIMILAR     571    752       WITH YEAST TOP1.
FT   CONFLICT    404    404       F -> L (IN REF. 1).
SQ   SEQUENCE   752 AA;  87828 MW;  CBD92432F099AD22 CRC64;


  Query Match             12.0%;  Score 75;  DB 1;  Length 752;
  Best Local Similarity   25.0%;  Pred. No. 35;
  Matches   27;  Conservative   21;  Mismatches   46;  Indels   14;  Gaps    5;

Qy      16 EEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKA 75
           | |:   |:|   |||  | |:|::  :   :  | :       | ::    |: :

Db     647 ERKKRALEEEASFPER--EQKIKSQRINSASQTEGDEL------KSEQTQPKGEISEENT 698

Qy      76 KMKNKQLPSAGPDKNLVTGDHIP--TPQDLPQRK---SSLVTSKLAGG 118
           |:|: :: |  ||   | |:  |  |   |   ::  || :: |
Db     699 KIKSSEVSSQDPDSG-VAGEFAPQNTTAQLENPKTEDNNAATSNISNG 745


RESULT  32
INCE_CHICK
ID   INCE_CHICK     STANDARD;      PRT;   877 AA.
AC   P53352;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Inner centromere protein (INCENP).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94012983; PubMed=8408220;
RA   Mackay A.M., Eckley D.M., Chue C., Earnshaw W.C.;
RT   "Molecular analysis of the INCENPs (inner centromere proteins):
RT   separate domains are required for association with microtubules
RT   during interphase and with the central spindle during anaphase.";
RL   J. Cell Biol. 123:373-385(1993).
CC   -!- FUNCTION: MAY ACT AS CYTOSKELETAL PROTEINS THAT ARE POTENTIALLY
CC       CAPABLE OF ALTERING THE MORPHOLOGY OF THE CELLULAR MICROTUBULE
CC       NETWORK DURING INTERPHASE.
CC   -!- SUBUNIT: HOMODIMER OR HETERODIMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR. IS RESTRICTED TO THE NUCLEUS IN
CC       INTERPHASE, REMAINS TIGHTLY BOUND TO THE CHROMOSOMES UNTIL EARLY
CC       METAPHASE, AND DURING LATE METAPHASE GETS CONCENTRATED IN LINEAR
CC       ARRAYS THAT TRANSECT THE METAPHASE PLATE BETWEEN THE CHROMOSOMES.
CC       AS ANAPHASE BEGINS TO MOVE TO THE SPINDLE MIDZONE WHERE IT IS
CC       INTIMATELY ASSOCIATED WITH THE BUNDLED MICROTUBULES. LATER IN
CC       ANAPHASE GETS CLOSELY ASSOCIATED WITH THE CELL CORTEX, AND BY
CC       TELOPHASE IS CONCENTRATED AT EACH SIDE OF THE MIDBODY IN THE
CC       INTERCELLULAR BRIDGE, WITH WHICH IT IS DISCARDED AFTER.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; INCENP I AND INCENP II (SHOWN
CC       HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z25419; CAA80906.1; -.
DR   EMBL; Z25420; CAA80907.1; -.
KW   Cell division; Microtubules; Coiled coil; Centromere; Mitosis;
KW   Cell cycle; Nuclear protein; Alternative splicing.
FT   DOMAIN      503    733       COILED COIL (POTENTIAL).
FT   VARSPLIC    716    753       MISSING (IN ISOFORM INCENP I).
FT   VARIANT     257    257       A -> T.
FT   VARIANT     471    471       E -> Q.
SQ   SEQUENCE   877 AA;  100940 MW;  AFA703149F555352 CRC64;


  Query Match             12.0%;  Score 75;  DB 1;  Length 877;
  Best Local Similarity   20.6%;  Pred. No. 41;
  Matches   27;  Conservative   28;  Mismatches   54;  Indels   22;  Gaps    3;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGIL-----------PERAEEAKLKAKYPSLGQKP-- 48
           :|:||:  | :  |::|  ||:|  |             : || :::|:     |:
Db     635 KKEEEQERARKIAEQRQAEQEREKQLAAEREQERKKEQERKKEEERIQAEKQREQQEKAA 694

Qy      49 -------GGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQ 101
                  : | | ::| :|  :     | : : : |:||     |::    |:   :
```

```
Db    695 RLQKEVLAAKEQLQKEMEKKEKE-EQLLAEMKRQEQEQKKLPEEQKAKDVAQTQHLENKE 753

Qy    102 DLPQRKSSLVT 112
          : |   |  :|
Db    754 NSPACNSYQMT 764


RESULT  33
ANK2_HUMAN
ID   ANK2_HUMAN     STANDARD;      PRT;  3924 AA.
AC   Q01484; Q01485;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).
GN   ANK2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC   TISSUE=Brain stem;
RX   MEDLINE=91302466; PubMed=1830053;
RA   Otto E., Kunimoto M., McLaughlin T., Bennett V.;
RT   "Isolation and characterization of cDNAs encoding human brain
RT   ankyrins reveal a family of alternatively spliced genes.";
RL   J. Cell Biol. 114:241-253(1991).
RN   [2]
RP   REVISIONS.
RA   Carpenter S.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain stem;
RX   MEDLINE=94075409; PubMed=8253844;
RA   Chan W., Kordeli E., Bennett V.;
RT   "440-kD ankyrinB: structure of the major developmentally regulated
RT   domain and selective localization in unmyelinated axons.";
RL   J. Cell Biol. 123:1463-1473(1993).
RN   [4]
RP   SEQUENCE OF 463-495 FROM N.A.
RX   MEDLINE=92009921; PubMed=1833308;
RA   Tse W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,
RA   Lux S.E., Ward D.C., Forget B.G.;
RT   "Isolation and chromosomal localization of a novel nonerythroid
RT   ankyrin gene.";
RL   Genomics 10:858-866(1991).
CC   -!- FUNCTION: Attach integral membrane proteins to cytoskeletal
CC       elements. Also bind to cytoskeletal proteins.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
CC       CELLS THROUGHOUT THE BRAIN.
CC   -!- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
CC       AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE
CC       AND FUNCTION (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 23 ANK REPEATS.
CC   -!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56957; CAA40278.1; -.
DR   EMBL; X56958; CAA40279.2; -.
DR   EMBL; Z26634; CAB42644.1; -.
DR   EMBL; M37123; AAA62828.1; -.
DR   PIR; S14533; S14533.
DR   PIR; A39643; A39643.
DR   PIR; B39643; B39643.
DR   PIR; S14569; S14569.
DR   HSSP; Q00421; 1AWC.
DR   MIM; 106410; -.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000488; Death.
DR   InterPro; IPR000906; ZU5.
DR   Pfam; PF00023; ank; 24.
DR   Pfam; PF00531; death; 1.
DR   Pfam; PF00791; ZU5; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 21.
DR   SMART; SM00005; DEATH; 1.
DR   SMART; SM00218; ZU5; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 20.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50017; DEATH_DOMAIN; 1.
KW   Cytoskeleton; Alternative splicing; Repeat; ANK repeat;
KW   Phosphorylation.
FT   REPEAT       63     92       ANK 1.
FT   REPEAT       96    125       ANK 2.
FT   REPEAT      129    158       ANK 3.
FT   REPEAT      162    191       ANK 4.
FT   REPEAT      193    220       ANK 5.
FT   REPEAT      232    261       ANK 6.
FT   REPEAT      265    294       ANK 7.
FT   REPEAT      298    327       ANK 8.
FT   REPEAT      331    360       ANK 9.
FT   REPEAT      364    393       ANK 10.
FT   REPEAT      397    426       ANK 11.
FT   REPEAT      430    459       ANK 12.
FT   REPEAT      463    492       ANK 13.
FT   REPEAT      496    525       ANK 14.
FT   REPEAT      529    558       ANK 15.
FT   REPEAT      562    591       ANK 16.
FT   REPEAT      595    624       ANK 17.
FT   REPEAT      628    657       ANK 18.
FT   REPEAT      661    690       ANK 19.
FT   REPEAT      694    723       ANK 20.
FT   REPEAT      727    756       ANK 21.
FT   REPEAT      760    789       ANK 22.
FT   REPEAT      793    822       ANK 23.
FT   DOMAIN     1773   1950       REPEAT-RICH REGION.
FT   REPEAT     1773   1784       REPEAT A.
FT   REPEAT     1785   1796       REPEAT A.
FT   REPEAT     1797   1808       REPEAT A.
FT   REPEAT     1809   1820       REPEAT A.
FT   REPEAT     1821   1832       REPEAT A.
FT   REPEAT     1833   1844       REPEAT A.
FT   REPEAT     1845   1856       REPEAT A.
FT   REPEAT     1857   1867       REPEAT A (APPROXIMATE).
FT   REPEAT     1868   1879       REPEAT A.
FT   REPEAT     1880   1891       REPEAT A.
FT   REPEAT     1892   1902       REPEAT A (APPROXIMATE).
FT   REPEAT     1903   1914       REPEAT A.
FT   REPEAT     1915   1926       REPEAT A.
FT   REPEAT     1927   1938       REPEAT A.
FT   REPEAT     1939   1950       REPEAT A.
FT   DOMAIN     3536   3620       DEATH.
FT   VARSPLIC   1039   1039       Q -> QFLGKLHLPTAPPPLNEGESLVSRILQLGPPGTK
FT                                (IN ISOFORM 2).
FT   VARSPLIC   1444   3528       MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT   CONFLICT    475    476       GQ -> PE (IN REF. 4).
FT   CONFLICT    971    971       I -> S (IN REF. 1).
FT   CONFLICT   3581   3582       QY -> HA (IN REF. 1).
FT   CONFLICT   3586   3586       I -> Y (IN REF. 1).
SQ   SEQUENCE   3924 AA;  430337 MW;  52AC496C428E29D2 CRC64;


  Query Match             12.0%;  Score 75;  DB 1;  Length 3924;
  Best Local Similarity   30.5%;  Pred. No. 2.1e+02;
  Matches   39;  Conservative   21;  Mismatches   48;  Indels   20;  Gaps    8;
```

```
Qy       3 QKQEEEN---PAEETGEEKQDTQEKEGILPERAEE--AKLKAKYPSLGQKPGGS-----D 52
           |||:||     ||:   :|  ::|  |  |  | |    :|| | : : | ||
Db    1679 QKQKEEGLQASAEKAELKKGSSEESLGEDPGLAPEPLPTVKATSPLIEETPIGSIKDKVK 1738

Qy      53 FLMKRL---QKGQKYFD---SGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIP-TPQDLPQ 105
            | ||:   |||:        |  :: | |  ::  |:| |  :| :  | | :|    :
Db    1739 ALQKRVEDEQKGRSKLPIRVKGKEDVPK-KTTHRPHPAASP--SLKSERHAPGSPSPKTE 1795

Qy     106 RKSSLVTS 113
           | |:| :|
Db    1796 RHSTLSSS 1803


RESULT  34
DD21_MOUSE
ID   DD21_MOUSE     STANDARD;      PRT;   851 AA.
AC   Q9JIK5; Q9WV45;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Nucleolar RNA helicase II (Nucleolar RNA helicase Gu) (RH II/Gu)
DE   (DEAD-box protein 21).
GN   DDX21.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ;
RX   MEDLINE=20318620; PubMed=10860663;
RA   Valdez B.C., Wang W.;
RT   "Mouse RNA helicase II/Gu: cDNA and genomic sequences, chromosomal
RT   localization, and regulation of expression.";
RL   Genomics 66:184-194(2000).
RN   [2]
RP   REVISIONS OF N-TERMINUS.
RA   Valdez B.C.;
RL   Submitted (APR-2001) to the SWISS-PROT data bank.
CC   -!- FUNCTION: CAN UNWIND DOUBLE-STRANDED RNA (HELICASE) AND CAN FOLD
CC       OR INTRODUCE A SECONDARY STRUCTURE TO A SINGLE-STRANDED RNA
CC       (FOLDASE).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LIVER AND TESTIS.
CC       EXPRESSED AT LOWER LEVEL IN BRAIN, LUNGS, AND SKELETAL MUSCLE.
CC   -!- DOMAIN: THE TWO ENZYMATIC ACTIVITIES RESIDE IN TWO SEPARATE
CC       DOMAINS, THE HELICASE IN THE N-TERMINUS AND THE FOLDASE IN THE C-
CC       TERMINUS.
CC   -!- DOMAIN: THE 3 X 5 AA REPEATS SEEM TO BE CRITICAL FOR THE RNA
CC       FOLDING ACTIVITY.
CC   -!- DOMAIN: THE 3 X 37 AA TANDEM REPEATS MAY BE IMPORTANT FOR NUCLEAR
CC       LOCALIZATION AND/OR INTERACTION WITH NUCLEIC ACIDS OR OTHER
CC       PROTEINS.
CC   -!- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF220365; AAF61690.1; -.
DR   EMBL; AF159131; AAD43959.3; -.
DR   MGD; MGI:1860494; Ddx21.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR000629; DEAD_ATP_helicase.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00039; DEAD_ATP_HELICASE; FALSE_NEG.
KW   Helicase; RNA-binding; ATP-binding; Nuclear protein; Antigen; Repeat.
FT   NP_BIND     302    309       ATP (POTENTIAL).
FT   SITE        411    414       DEVD BOX.
FT   DOMAIN       95     98       POLY-LYS.
FT   DOMAIN      117    227       3 X 37 AA TANDEM REPEATS.
FT   REPEAT      117    153       1-1.
FT   REPEAT      154    190       1-2.
FT   REPEAT      191    227       1-3.
FT   DOMAIN      807    833       3 X 5 AA REPEATS.
FT   REPEAT      807    811       2-1.
FT   REPEAT      817    823       2-2.
FT   REPEAT      829    833       2-3.
FT   CONFLICT      1     46       MISSING (IN REF. 1; AAF61690).
FT   CONFLICT    790    790       Q -> P (IN REF. 1; AAF61690).
SQ   SEQUENCE   851 AA;  93582 MW;  71D7286F4CA1DDD7 CRC64;


  Query Match             11.9%;  Score 74.5;  DB 1;  Length 851;
  Best Local Similarity   25.2%;  Pred. No. 44;
  Matches   30;  Conservative   22;  Mismatches   32;  Indels   35;  Gaps    6;

Qy       4 KQEEENPAEETG------------EEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGG 50
           |::|| | ::|               |:: |:: || |  | :||   :|  |    ||
Db      77 KKQEEEPQDDTASTSKTSKKKKEPLEKQADSETKEIITEEPSEE---EADMP----KP-- 127

Qy      51 SDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSS 109
                 |:::||::   :||      |:||      |  :       || |: : : | :
Db     128 -----KKMKKGKE--ANGDAGEKSPKLKN------GLSQPSEEEADIPKPKKMKKGKEA 173


RESULT  35
CHD2_HUMAN
ID   CHD2_HUMAN     STANDARD;      PRT;  1739 AA.
AC   O14647;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Chromodomain-helicase-DNA-binding protein 2 (CHD-2).
GN   CHD2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97470991; PubMed=9326634;
RA   Woodage T., Basrai M.A., Baxevanis A.D., Hieter P., Collins F.S.;
RT   "Characterization of the CHD family of proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997).
CC   -!- FUNCTION: SEQUENCE-SELECTIVE DNA-BINDING PROTEIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 'CHROMO' DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF006514; AAB87382.1; -.
DR   HSSP; P23197; 1AP0.
DR   MIM; 602119; -.
DR   InterPro; IPR000953; Chromo.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00385; chromo; 2.
DR   Pfam; PF00271; helicase_C; 1.
```

```
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00298; CHROMO; 2.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00598; CHROMO_1; 2.
DR   PROSITE; PS50013; CHROMO_2; 2.
KW   DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat.
FT   DOMAIN        1     74       SER-RICH.
FT   DOMAIN      121    139       SER-RICH.
FT   DOMAIN      261    353       CHROMO 1.
FT   DOMAIN      378    456       CHROMO 2.
FT   NP_BIND     509    516       ATP (POTENTIAL).
FT   SITE        617    620       DEAH BOX.
SQ   SEQUENCE   1739 AA;  200560 MW;  D62089C25EBAA40F CRC64;


  Query Match             11.9%;  Score 74.5;  DB 1;  Length 1739;
  Best Local Similarity   29.3%;  Pred. No. 96;
  Matches   36;  Conservative   17;  Mismatches   53;  Indels   17;  Gaps    7;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGI--LPERAEEAKLKAKYPSLGQKPGGSDFLMKRL 58
           :|  :  :||:|| || | |  ||  :    :: |  : | |  |  :   |     |
Db    1363 LSSPRHSDNPSEE-GEVKDDGLEKSPMKKKQKKKENKENKEKQMSSRKDKEGDKERKKSK 1421

Qy      59 QKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTP------QDLPQRKSSLVT 112
             | :|   |||   ||:  |:|:  | ||    :|    | |       || |   |:
Db    1422 DKKEKP-KSGD---AKSSSKSKR--SQGPVH--ITAGSEPVPIGEDEDDDLDQETFSICK 1473

Qy     113 SKL 115
           ::
Db    1474 ERM 1476


RESULT  36
PRKD_HUMAN
ID   PRKD_HUMAN     STANDARD;      PRT;  4128 AA.
AC   P78527; Q14175; P78528; Q13337; Q13327;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA-dependent protein kinase catalytic subunit (EC 2.7.1.37) (DNA-
DE   PKcs) (DNPK1).
GN   PRKDC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cervical carcinoma;
RX   MEDLINE=95401275; PubMed=7671312;
RA   Hartley K.O., Gell D., Smith G.C.M., Zhang H., Divecha N.,
RA   Connelly M.A., Admon A., Lees-Miller S.P., Anderson C.W.,
RA   Jackson S.P.;
RT   "DNA-dependent protein kinase catalytic subunit: a relative of
RT   phosphatidylinositol 3-kinase and the ataxia telangiectasia gene
RT   product.";
RL   Cell 82:849-856(1995).
RN   [2]
RP   REVISIONS.
RA   Gell D., Anderson C.W.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-1689 FROM N.A.
RA   Anderson C.W., Freimuth P.I., Dunn J.J.;
RT   "Partial sequence of a BAC insert containing exons 8 through 38 of the
RT   human gene encoding DNA-PKcs.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1789-2203 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=95365397; PubMed=7638222;
RA   Sipley J.D., Menninger J.C., Hartley K.O., Ward D.C., Jackson S.P.,
RA   Anderson C.W.;
RT   "Gene for the catalytic subunit of the human DNA-activated protein
RT   kinase maps to the site of the XRCC7 gene on chromosome 8.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:7515-7519(1995).
RN   [5]
RP   SEQUENCE OF 3199-4128 FROM N.A.
RC   TISSUE=Fetal lung;
RX   MEDLINE=96062265; PubMed=7594449;
RA   Poltoratsky V.P., Shi X., York J.D., Lieber M.R., Carter T.H.;
RT   "Human DNA-activated protein kinase (DNA-PK) is homologous to
RT   phosphatidylinositol kinases.";
RL   J. Immunol. 155:4529-4533(1995).
RN   [6]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=97074683; PubMed=8917110;
RA   Connelly M.A., Zhang H., Kieleczawa J., Anderson C.W.;
RT   "Alternate splice-site utilization in the gene for the catalytic
RT   subunit of the DNA-activated protein kinase, DNA-PKcs.";
RL   Gene 175:271-273(1996).
RN   [7]
RP   CHARACTERIZATION.
RX   MEDLINE=98040126; PubMed=9372844;
RA   Wu X., Lieber M.R.;
RT   "Interaction between DNA-dependent protein kinase and a novel protein,
RT   KIP.";
RL   Mutat. Res. 385:13-20(1997).
CC   -!- FUNCTION: SER/THR KINASE INVOLVED IN DNA DOUBLE-STRANDED BREAK
CC       REPAIR, V(D)J RECOMBINATION AND MODULATION OF TRANSCRIPTION. MUST
CC       BE BOUND TO DNA TO EXPRESS ITS CATALYTIC PROPERTIES.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SUBUNIT: FORMS A COMPLEX WITH THE HETERODIMER KU70/KU80
CC       AUTOANTIGEN. INTERACTS WITH DNA-PKCS INTERACTING PROTEIN (KIP)
CC       WITH THE REGION UPSTREAM THE KINASE DOMAIN.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U47077; AAB39925.3; -.
DR   EMBL; U34994; AAC50210.2; -.
DR   EMBL; U63630; AAC52019.2; -.
DR   EMBL; U35835; AAA79184.1; -.
DR   EMBL; L27425; AAA79244.1; -.
DR   MIM; 600899; -.
DR   InterPro; IPR003151; FAT.
DR   InterPro; IPR003152; FATC.
DR   InterPro; IPR000403; PI3_PI4_kinase.
DR   Pfam; PF02259; FAT; 1.
DR   Pfam; PF02260; FATC; 1.
DR   Pfam; PF00454; PI3_PI4_kinase; 1.
DR   SMART; SM00146; PI3Kc; 1.
DR   PROSITE; PS00915; PI3_4_KINASE_1; 1.
DR   PROSITE; PS00916; PI3_4_KINASE_2; FALSE_NEG.
DR   PROSITE; PS50290; PI3_4_KINASE_3; 1.
KW   Transferase; Serine/threonine-protein kinase; DNA repair;
KW   Nuclear protein; Alternative splicing.
FT   DOMAIN     2436   3212       KIP BINDING.
FT   DOMAIN     3748   4128       PI3K/PI4K.
FT   VARSPLIC   3799   3829       MISSING (IN ISOFORM 2).
FT   CONFLICT    405    405       Y -> D (IN REF. 3).
FT   CONFLICT   1008   1008       S -> A (IN REF. 3).
FT   CONFLICT   3434   3434       T -> I (IN REF. 5).
FT   CONFLICT   3660   3660       N -> T (IN REF. 5).
FT   CONFLICT   3817   3817       L -> W (IN REF. 5).
FT   CONFLICT   3862   3862       A -> P (IN REF. 5).
```

```
SQ   SEQUENCE   4128 AA;  469136 MW;  23E1861CE68606F4 CRC64;


  Query Match           11.9%;  Score 74.5;  DB 1;  Length 4128;
  Best Local Similarity   25.7%;  Pred. No. 2.4e+02;
  Matches   27;  Conservative   14;  Mismatches   37;  Indels   27;  Gaps    5;

Qy      20 DTQEKEGILPERAEEAKLKAKYPSLGQ---KPGGSDFLMKRLQKGQKYFD---------- 66
           :||  :| |  | :|  | |::|  ||        || : :   |:  ||
Db    2608 ETQASQGTLQTRTQEGSLSARWPVAGQIRATQQQHDFTLTQTADGRSSFDWLTGSSTDPL 2667

Qy      67 ------SGD----YNMAKAKMKNKQLPSAGPD---KNL-VTGDHI 97
                | |     :    :::   | | |||   | | : || :
Db    2668 VDHTSPSSDSLLFAHKRSERLQRAPLKSVGPDFGKKRLGLPGDEV 2712


RESULT  37
RDGB_DROME
ID   RDGB_DROME     STANDARD;      PRT;  1054 AA.
AC   P43125;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Retinal degeneration B protein (Probable calcium transporter RDGB).
GN   RDGB.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R; TISSUE=Head;
RX   MEDLINE=91231170; PubMed=1903119;
RA   Vihtelic T.S., Hyde D.R., O'Tousa J.E.;
RT   "Isolation and characterization of the Drosophila retinal
RT   degeneration B (rdgB) gene.";
RL   Genetics 127:761-768(1991).
CC   -!- FUNCTION: RDGB MUTANTS UNDERGO RAPID LIGHT-INDUCED RETINAL
CC       DEGENERATION. MAY CONTROL PHOSPHATIDYLINOSITOL CONCENTRATION IN
CC       TRANSPORT VESICLES FROM THE SUBRHABDOMERIC CISTERNAE (SRC) TO THE
CC       RHABDOMERE. MAY FUNCTION AS A CALCIUM TRANSPORTER.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN ADULT HEADS, NOT DETECTED IN
CC       BODIES.
CC   -!- SIMILARITY: THE N-TERMINAL IS SIMILAR TO MAMMALIAN
CC       PHOSPHATIDYLINOSITOL (PTDINS) TRANSFER PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57978; CAA41044.1; -.
DR   PIR; A61221; A61221.
DR   FlyBase; FBgn0003218; rdgB.
DR   InterPro; IPR004177; DDHD.
DR   InterPro; IPR001666; PI_transfer.
DR   Pfam; PF02862; DDHD; 1.
DR   Pfam; PF02121; IP_trans; 1.
DR   PRINTS; PR00391; PITRANSFER.
KW   Transmembrane; Vision.
FT   DOMAIN      321    332       ASP/GLU-RICH (HIGHLY ACIDIC).
FT   TRANSMEM    501    519       POTENTIAL.
FT   TRANSMEM    734    750       POTENTIAL.
FT   CARBOHYD    194    194       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    414    414       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    612    612       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    658    658       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    852    852       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    928    928       N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ   SEQUENCE   1054 AA;  116560 MW;  87F26EB871003CA8 CRC64;


  Query Match           11.8%;  Score 73.5;  DB 1;  Length 1054;
  Best Local Similarity   23.8%;  Pred. No. 68;
  Matches   34;  Conservative   18;  Mismatches   50;  Indels   41;  Gaps    6;

Qy       3 QKQEEENPAEETGEEKQDTQEKE-------------GILPERAEEAKL----KAKYPSL 44
           |    |  ::| |||::|  | |              |   :|:    :    | |: |
Db     311 QPPSAEASSDEEGEEEEDDDEDENDAIGTGVDLSANQGGSAQRSRSQSIQMAQKGKFGSK 370

Qy      45 G--QKPGGS---------DFLMKRLQ------KGQKYFDSGDYNMAKAKMKNKQLPSAGP 87
           |    | ||         :: |:||:        :::||  | |   :  |   |   |
Db     371 GALHSPVGSAHSFDLQVANWRMERLEVDSKSNSDEEFFDCLDTNETNSLAKWSSLELLG- 429

Qy      88 DKNLVTGDHIPTPQDLPQRKSSL 110
                ||  | |   |   :|:
Db     430 -----EGDDSPPPHGGPSSAASV 447


RESULT  38
HRX_MOUSE
ID   HRX_MOUSE      STANDARD;      PRT;  3866 AA.
AC   P55200;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Zinc finger protein HRX (ALL-1) (Fragment).
GN   MLL OR HRX OR ALL1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   STRAIN=C57BL/6J, AND B6/CBA; TISSUE=Spleen, and Lung;
RX   MEDLINE=93317679; PubMed=8327517;
RA   Ma Q., Alder H., Nelson K.K., Chatterjee D., Gu Y., Nakamura T.,
RA   Canaani E., Croce C.M., Siracusa L.D., Buchberg A.M.;
RT   "Analysis of the murine All-1 gene reveals conserved domains with
RT   human ALL-1 and identifies a motif shared with DNA
RT   methyltransferases.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:6350-6354(1993).
CC   -!- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
CC       MAY REGULATE GENES INVOLVED IN SKELETAL FORMATION DURING
CC       EMBRYONIC DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
CC   -!- SIMILARITY: CONTAINS 1 CXXC-TYPE ZINC FINGER.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L17069; AAA62593.1; -.
DR   MGD; MGI:96995; Mll.
DR   InterPro; IPR001487; Bromodomain.
DR   InterPro; IPR003889; FYrich_C.
DR   InterPro; IPR003888; FYrich_N.
DR   InterPro; IPR001965; PHD.
DR   InterPro; IPR003616; PostSET.
DR   InterPro; IPR000093; RecR.
DR   InterPro; IPR001214; SET.
DR   InterPro; IPR002857; Znf-CXXC.
```

```
DR   Pfam; PF00628; PHD; 3.
DR   Pfam; PF00856; SET; 1.
DR   Pfam; PF02008; zf-CXXC; 1.
DR   SMART; SM00297; BROMO; 1.
DR   SMART; SM00542; FYRC; 1.
DR   SMART; SM00541; FYRN; 1.
DR   SMART; SM00249; PHD; 4.
DR   SMART; SM00508; PostSET; 1.
DR   SMART; SM00317; SET; 1.
DR   PROSITE; PS50014; BROMODOMAIN_2; 1.
DR   PROSITE; PS50280; SET; 1.
KW   DNA-binding; Nuclear protein; Zinc-finger; Metal-binding;
KW   Transcription regulation; Alternative splicing; Polymorphism.
FT   NON_TER       1      1
FT   DNA_BIND     67     78       A.T HOOK (BY SIMILARITY).
FT   DNA_BIND    115    125       A.T HOOK (BY SIMILARITY).
FT   DNA_BIND    199    207       A.T HOOK (BY SIMILARITY).
FT   ZN_FING    1044   1091       CXXC-TYPE.
FT   ZN_FING    1330   1381       PHD-TYPE 1.
FT   ZN_FING    1383   1432       PHD-TYPE 2.
FT   ZN_FING    1465   1529       PHD-TYPE 3.
FT   DOMAIN     1605   1650       BROMODOMAIN (DIVERGENT).
FT   DOMAIN     3737   3846       SET.
FT   DOMAIN       35     41       POLY-GLY.
FT   DOMAIN      459    469       PRO-RICH.
FT   DOMAIN     1231   1238       POLY-PRO.
FT   DOMAIN     3533   3536       POLY-GLU.
FT   DOMAIN     3693   3697       POLY-GLU.
FT   VARSPLIC   1503   1505       MISSING (IN ISOFORM 2).
FT   VARIANT    1497   1497       K -> T.
SQ   SEQUENCE   3866 AA;  420976 MW;  ADFC55E14E806F1D CRC64;
```

```
  Query Match             11.8%;  Score 73.5;  DB 1;  Length 3866;
  Best Local Similarity   25.2%;  Pred. No. 2.8e+02;
  Matches   31;  Conservative   18;  Mismatches   53;  Indels   21;  Gaps    6;

Qy      13 ETGEEKQDTQEKEGILPERAEEAKLKA-KYPSLGQKP-------GGSDFLMKRLQKGQKY 64
           :  :  |     : : |   |: ::|  | | :| :|       | |     | |||:|
Db    2273 QRSDRDQHMDPSQSVKPSPNEDGEIKTLKLPGMGHRPSILHEHIGSSS--RDRRQKGKK- 2329

Qy      65 FDSGDYNMAKAKMKNKQL-----PSAGPDKNLVT--GDHIPTPQDLPQRKSSLVTSKLAG 117
                 | |  :|       : | : ||     | : ||  | ||  : |:|:  |
Db    2330 ---SSKETCKEKHSSKSYLEPGQVTTGEEGNLKPEFADEVLTPGFLGQRPCNNVSSEKIG 2386

Qy     118 GQV 120
             :|
Db    2387 DKV 2389
```

```
RESULT  39
YBB9_SCHPO
ID   YBB9_SCHPO     STANDARD;      PRT;   449 AA.
AC   O60071;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 51.4 kDa protein C13G1.09 in chromosome II.
GN   SPBC13G1.09.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Lyne M., Wood V., Rajandream M.A., Barrell B.G., Badcock K.,
RA   Churcher C.M.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE BYSTIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL022600; CAA18662.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   449 AA;  51365 MW;  06F4C32CFEBF3C3B CRC64;
```

```
  Query Match             11.7%;  Score 73;  DB 1;  Length 449;
  Best Local Similarity   20.8%;  Pred. No. 30;
  Matches   30;  Conservative   28;  Mismatches   44;  Indels   42;  Gaps    5;

Qy       3 QKQEEENPAEETGEEK--------------QDTQEKEGILPERAEEAKLKAKYPSLGQK 47
           ||   |||| |:  |:               |: :| | :  : |:        |:: :
Db      86 QKDTLENPAIESSYEESEHARNVSDSESITSQEEEEYEELEIDDADRDLFDRFLPTVSGE 145

Qy      48 PGG---------SDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIP 98
             |           || :|:::             | |:|: : : :|||  ::| :     |
Db     146 EGDLTEEKTTSLSDLIMQKI------------NEAEARARGEYIPSAEEEENALP----P 189

Qy      99 TPQDLPQ--RKSSLVTSKLAGGQV 120
             |  : :   |  :: ||   |::
Db     190 LPPKVIEVYSKVGVLLSKYRSGKI 213
```

```
RESULT  40
HEX_ADE01
ID   HEX_ADE01      STANDARD;      PRT;   467 AA.
AC   Q04965;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hexon protein (Late protein 2) (Fragment).
GN   PII.
OS   Human adenovirus type 1.
OC   Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX   NCBI_TaxID=10533;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93289026; PubMed=8511397;
RA   Pring-Akerblom P., Adrian T.;
RT   "The hexon genes of adenoviruses of subgenus C: comparison of the
RT   variable regions.";
RL   Res. Virol. 144:117-127(1993).
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF THE STRUCTURAL PROTEINS IN THE
CC       VIRAL COAT AND IS SYNTHESIZED DURING LATE INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67709; CAA47946.1; -.
DR   HSSP; P03277; 1DHX.
DR   InterPro; IPR000736; Adeno_hexon.
DR   Pfam; PF01065; Adeno_hexon; 1.
DR   ProDom; PD002815; Adeno_hexon; 1.
KW   Coat protein; Hexon protein; Late protein.
FT   NON_TER       1      1
FT   DOMAIN       15     45       GLU-RICH (ACIDIC).
FT   NON_TER     467    467
SQ   SEQUENCE   467 AA;  52060 MW;  DC72E5BBFA6FB825 CRC64;
```

```
  Query Match             11.7%;  Score 73;  DB 1;  Length 467;
```

```
  Best Local Similarity   26.9%;  Pred. No. 31;
  Matches   29;  Conservative   16;  Mismatches   29;  Indels   34;  Gaps    5;

Qy       6 EEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           |:| | :|   | :| :| |       |||: :|: |  |||          ::|   |
Db      17 EQEEPTQEMAGELEDEEEAE------EEEAEEEAEAPQAGQK----------VKKTHVY- 59

Qy      66 DSGDYNMAKAKMKNKQLPSAG--------PDKNLVTGDHIPTPQDLPQ 105
                 |:| :  ::: : |         : | |  |  || |  ||
Db      60 -------AQAPLAGEKITANGLQIVSDTQTEGNPVFAD--PTYQPEPQ 98


Search completed: September  7, 2002, 10:51:19
Job time: 395 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:43:49 ; Search time 20.63 Seconds
(without alignments)
143.262 Million cell updates/sec

Title: US-09-824-178-4
Perfect score: 625
Sequence: 1 MSQKQEEENPAEETGEEKQD.....DLPQRKSSLVTSKLAGGQVE 121

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	471	75.4	118	2	US-08-739-484-1	Sequence 1, Appli
2	425.5	68.1	112	2	US-08-739-484-3	Sequence 3, Appli
3	414	66.2	96	2	US-08-739-484-4	Sequence 4, Appli
4	83	13.3	619	1	US-08-465-746-2	Sequence 2, Appli
5	83	13.3	619	1	US-08-214-164-2	Sequence 2, Appli
6	83	13.3	619	2	US-08-467-852A-3	Sequence 3, Appli
7	83	13.3	619	2	US-08-246-636-2	Sequence 2, Appli
8	83	13.3	619	2	US-08-247-491A-3	Sequence 3, Appli
9	83	13.3	619	2	US-08-319-795-2	Sequence 2, Appli
10	83	13.3	619	2	US-08-468-985-2	Sequence 2, Appli
11	83	13.3	619	3	US-08-312-949-2	Sequence 2, Appli
12	83	13.3	648	1	US-08-072-070-2	Sequence 2, Appli
13	83	13.3	648	1	US-08-469-434-2	Sequence 2, Appli
14	83	13.3	648	1	US-08-214-222-2	Sequence 2, Appli
15	83	13.3	648	2	US-08-467-852A-2	Sequence 2, Appli
16	83	13.3	648	2	US-08-468-718-2	Sequence 2, Appli
17	83	13.3	648	2	US-08-247-491A-2	Sequence 2, Appli
18	83	13.3	648	3	US-08-446-201-3	Sequence 3, Appli
19	83	13.3	695	1	US-08-127-499A-23	Sequence 23, Appl
20	83	13.3	695	1	US-08-482-847-23	Sequence 23, Appl
21	81.5	13.0	473	1	US-08-363-215-2	Sequence 2, Appli
22	81.5	13.0	473	3	US-08-807-342B-7	Sequence 7, Appli
23	81.5	13.0	473	4	US-08-664-962B-4	Sequence 4, Appli
24	81.5	13.0	473	4	US-09-311-743-4	Sequence 4, Appli
25	81	13.0	423	1	US-08-353-550-3	Sequence 3, Appli
26	81	13.0	423	2	US-08-551-687-3	Sequence 3, Appli
27	81	13.0	474	1	US-08-353-550-8	Sequence 8, Appli
28	81	13.0	474	2	US-08-551-687-8	Sequence 8, Appli
29	79.5	12.7	479	4	US-09-446-504-64	Sequence 64, Appl
30	79.5	12.7	479	4	US-09-712-266-64	Sequence 64, Appl
31	78	12.5	411	2	US-08-741-134-6	Sequence 6, Appli
32	76	12.2	629	3	US-08-556-419-23	Sequence 23, Appl
33	75.5	12.1	1027	4	US-08-446-137B-2	Sequence 2, Appli
34	75	12.0	344	6	5210183-2	Patent No. 5210183
35	75	12.0	683	6	5210183-3	Patent No. 5210183
36	73.5	11.8	1250	3	US-08-938-291A-9	Sequence 9, Appli
37	72.5	11.6	1426	4	US-09-136-574A-43	Sequence 43, Appl
38	72	11.5	312	1	US-08-285-440-4	Sequence 4, Appli
39	72	11.5	312	1	US-08-630-349-4	Sequence 4, Appli
40	72	11.5	532	1	US-08-285-440-5	Sequence 5, Appli
41	72	11.5	532	1	US-08-630-349-5	Sequence 5, Appli
42	72	11.5	558	1	US-08-285-440-6	Sequence 6, Appli
43	72	11.5	558	1	US-08-630-349-6	Sequence 6, Appli
44	72	11.5	564	4	US-09-308-022-6	Sequence 6, Appli
45	72	11.5	631	4	US-08-847-065-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-08-739-484-1
; Sequence 1, Application US/08739484
; Patent No. 5917028
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN
; TITLE OF INVENTION: PHOSPHOPROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/739,484
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0140 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 118 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Consensus
US-08-739-484-1

Query Match 75.4%; Score 471; DB 2; Length 118;
Best Local Similarity 100.0%; Pred. No. 3.1e-44;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      31 RAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKN 90
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 RAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKN 87

Qy      91 LVTGDHIPTPQDLPQRKSSLVTSKLAGGQVE 121
           ||||||||||||||||||||||||||||||||
Db      88 LVTGDHIPTPQDLPQRKSSLVTSKLAGGQVE 118


RESULT   2
US-08-739-484-3
; Sequence 3, Application US/08739484
; Patent No. 5917028
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    TITLE OF INVENTION:  A NOVEL HUMAN
;    TITLE OF INVENTION:  PHOSPHOPROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/739,484
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0140 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  112 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  162690
US-08-739-484-3


  Query Match             68.1%;  Score 425.5;  DB 2;  Length 112;
  Best Local Similarity   76.4%;  Pred. No. 2.6e-39;
  Matches   84;  Conservative    9;  Mismatches   16;  Indels    1;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDS 67
           | |    : ||::: ::|    ||:|||||||:|| ||||||||||| ||||||||||
Db       4 EVPEAASAEEQKEMEDKV-TSPEKAEEAKLKARYPHLGQKPGGSDFLRKRLQKGQKYFDS 62

Qy      68 GDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           ||||||||||||||:| |||  |||||||||||||||| ||| |||||
Db      63 GDYNMAKAKMKNKQLPTATPDKTEVTGDHIPTPQDLPQRKPSLVASKLAG 112


RESULT   3
US-08-739-484-4
; Sequence 4, Application US/08739484
; Patent No. 5917028
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    TITLE OF INVENTION:  A NOVEL HUMAN
;    TITLE OF INVENTION:  PHOSPHOPROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/739,484
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0140 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  96 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  741603
US-08-739-484-4


  Query Match             66.2%;  Score 414;  DB 2;  Length 96;
  Best Local Similarity   88.8%;  Pred. No. 3.9e-38;
  Matches   79;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      29 PERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPD 88
           ||:|||||||:|| |||||||||| |||||||||||||||||||||||||:| ||
Db       8 PEKAEEAKLKARYPHLGQKPGGSDFLRKRLQKGQKYFDSGDYNMAKAKMKNKQLPTAAPD 67

Qy      89 KNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           | ||||||||||||||| ||| |||||
Db      68 KTEVTGDHIPTPQDLPQRKPSLVASKLAG 96


RESULT   4
US-08-465-746-2
; Sequence 2, Application US/08465746
; Patent No. 5679768
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E
;    APPLICANT:  Yother, Janet L
;    APPLICANT:  McDaniel, Larry S
;    TITLE OF INVENTION:  EPITOPIC REGIONS OF PNEUMOCOCCAL SURFACE
;    TITLE OF INVENTION:  PROTEIN A
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  Suite 1203, 2001 Jefferson Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/465,746
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/048,896
;      FILING DATE:
;      APPLICATION NUMBER:  US 07/656,773
;      FILING DATE:  15-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/835,698
;      FILING DATE:  12-FEB-1992
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 521-0378
;      TELEX:  LUKPAT WASHINGTON
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  619 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-465-746-2


  Query Match             13.3%;  Score 83;  DB 1;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 0.53;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT   5
US-08-214-164-2
; Sequence 2, Application US/08214164
; Patent No. 5728387
;  GENERAL INFORMATION:
;    APPLICANT:  BRILES, DAVID E.
;    APPLICANT:  YOTHER, JANET L.
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  Suite 1203, 2001 Jefferson Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/214,164
;      FILING DATE:  17-MAR-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/656,773
;      FILING DATE:  15-FEB-1991
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Berkstresser, Jerry W.
;      REGISTRATION NUMBER:  22,651
;      REFERENCE/DOCKET NUMBER:  6102-137
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 521-0813
;      TELEX:  LUKPAT WASHINGTON
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  619 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-214-164-2


  Query Match             13.3%;  Score 83;  DB 1;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 0.53;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT   6
US-08-467-852A-3
; Sequence 3, Application US/08467852A
; Patent No. 5856170
;  GENERAL INFORMATION:
;    APPLICANT:  BRILES, David E.
;    APPLICANT:  YOTHER, Janet L.
;    APPLICANT:  MCDANIEL, Larry S.
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;      STREET:  745 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10151
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/467,852A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOWALSKI, Thomas J.
;      REGISTRATION NUMBER:  32,147
;      REFERENCE/DOCKET NUMBER:  454312-2064
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-588-0800
;      TELEFAX:  212-588-0500
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  619 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  n/a
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  amino acid
US-08-467-852A-3


  Query Match             13.3%;  Score 83;  DB 2;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 0.53;
```

```
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
            : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT   7
US-08-246-636-2
; Sequence 2, Application US/08246636
; Patent No. 5965141
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E
;    APPLICANT:  Yother, Janet L
;    APPLICANT:  McDaniel, Larry S
;    APPLICANT:  Wu, Hong-Yin
;    TITLE OF INVENTION:  EPITOPIC REGIONS OF PNEUMOCOCCAL SURFACE
;    TITLE OF INVENTION:  PROTEIN A
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  Suite 1203, 2001 Jefferson Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/246,636
;      FILING DATE:  20-MAY-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/656,773
;      FILING DATE:  15-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/835,698
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/048,896
;      FILING DATE:  20-APR-1993
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;      TELEX:  LUKPAT WASHINGTON
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  619 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-246-636-2


  Query Match             13.3%;  Score 83;  DB 2;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 0.53;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
            : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT   8
US-08-247-491A-3
; Sequence 3, Application US/08247491A
; Patent No. 5965400
;  GENERAL INFORMATION:
;    APPLICANT:  BRILES, David E.
;    APPLICANT:  YOTHER, Janet L.
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;      STREET:  745 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10151
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/247,491A
;      FILING DATE:  23-JUN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOWALSKI, Thomas J.
;      REGISTRATION NUMBER:  32,147
;      REFERENCE/DOCKET NUMBER:  454312-2041
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-588-0800
;      TELEFAX:  212-588-0500
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  619 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  n/a
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  amino acid
US-08-247-491A-3


  Query Match             13.3%;  Score 83;  DB 2;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 0.53;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
            : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT   9
US-08-319-795-2
; Sequence 2, Application US/08319795
; Patent No. 5980909
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E.
;    APPLICANT:  Yother, Janet L.
;    APPLICANT:  McDaniel, Larry S
;    TITLE OF INVENTION:  Epitopic Regions of Pneumococcal Surface
;    TITLE OF INVENTION:  Protein A
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheomaker and Mattare, Ltd.
;      STREET:  1203 Crystal Plaza Bldg. 1, 2001 Jefferson
;      STREET:  Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
```

```
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/319,795
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/246,636
;      FILING DATE:  20-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/048,896
;      FILING DATE:  20-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/835,698
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/656,773
;      FILING DATE:  15-FEB-1991
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  619 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-319-795-2


  Query Match             13.3%;  Score 83;  DB 2;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 0.53;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  10
US-08-468-985-2
; Sequence 2, Application US/08468985
; Patent No. 5997882
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E.
;    APPLICANT:  Yother, Janet L.
;    APPLICANT:  McDaniel, Larry S
;    TITLE OF INVENTION:  Epitopic Regions of Pneumococcal Surface
;    TITLE OF INVENTION:  Protein A
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheomaker and Mattare, Ltd.
;      STREET:  1203 Crystal Plaza Bldg. 1, 2001 Jefferson
;      STREET:  Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/468,985
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/319,795
;      FILING DATE:
;      APPLICATION NUMBER:  US 08/246,636
;      FILING DATE:  20-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/048,896
;      FILING DATE:  20-APR-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/835,698
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/656,773
;      FILING DATE:  15-FEB-1991
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  619 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-468-985-2


  Query Match             13.3%;  Score 83;  DB 2;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 0.53;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  11
US-08-312-949-2
; Sequence 2, Application US/08312949
; Patent No. 6027734
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E.
;    APPLICANT:  Wu, Hong-Yin
;    TITLE OF INVENTION:  MUCOSAL ADMINISTRATION OF
;    TITLE OF INVENTION:  PNEUMOCOCCAL ANTIGENS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford, P.C.
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/312,949
;      FILING DATE:  30-SEP-1994
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454312-2049
;    TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  (212) 840-3333
;       TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  619 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-312-949-2


  Query Match             13.3%;  Score 83;  DB 3;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 0.53;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  12
US-08-072-070-2
; Sequence 2, Application US/08072070
; Patent No. 5476929
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E
;    APPLICANT:  Yother, Janet L
;    APPLICANT:  McDaniel, Larry S
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  Suite 1203, 2001 Jefferson Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/072,070
;      FILING DATE:  19930603
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/835,698
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/656,773
;      FILING DATE:  15-FEB-1991
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 521-0378
;      TELEX:  LUKPAT WASHINGTON
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  648 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-072-070-2


  Query Match             13.3%;  Score 83;  DB 1;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 0.56;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  13
US-08-469-434-2
; Sequence 2, Application US/08469434
; Patent No. 5753463
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E
;    APPLICANT:  Yother, Janet L
;    APPLICANT:  McDaniel, Larry S
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  Suite 1203, 2001 Jefferson Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,434
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/072,065
;      FILING DATE:  03 JUNE 1993
;      APPLICATION NUMBER:  US/07/835,698
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/656,773
;      FILING DATE:  15-FEB-1991
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 521-0378
;      TELEX:  LUKPAT WASHINGTON
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  648 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-469-434-2


  Query Match             13.3%;  Score 83;  DB 1;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 0.56;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  14
```

```
US-08-214-222-2
; Sequence 2, Application US/08214222
; Patent No. 5804193
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E
;    APPLICANT:  Yother, Janet L
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  Suite 1203, 2001 Jefferson Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/214,222
;      FILING DATE:  17-MAR-1994
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/835,698
;      FILING DATE:  12-FEB-1992
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 521-0378
;      TELEX:  LUKPAT WASHINGTON
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  648 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-214-222-2


  Query Match             13.3%;  Score 83;  DB 1;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 0.56;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  15
US-08-467-852A-2
; Sequence 2, Application US/08467852A
; Patent No. 5856170
;  GENERAL INFORMATION:
;    APPLICANT:  BRILES, David E.
;    APPLICANT:  YOTHER, Janet L.
;    APPLICANT:  MCDANIEL, Larry S.
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;      STREET:  745 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10151
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/467,852A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOWALSKI, Thomas J.
;      REGISTRATION NUMBER:  32,147
;      REFERENCE/DOCKET NUMBER:  454312-2064
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-588-0800
;      TELEFAX:  212-588-0500
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  648 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-467-852A-2


  Query Match             13.3%;  Score 83;  DB 2;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 0.56;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  16
US-08-468-718-2
; Sequence 2, Application US/08468718
; Patent No. 5871943
;  GENERAL INFORMATION:
;    APPLICANT:  Briles, David E
;    APPLICANT:  Yother, Janet L
;    APPLICANT:  McDaniel, Larry S
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  Suite 1203, 2001 Jefferson Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/468,718
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/072,068
;      FILING DATE:  03 JUNE 1993
;      APPLICATION NUMBER:  US/07/835,698
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/656,773
;      FILING DATE:  15-FEB-1991
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
```

```
;      TELEFAX:  (703) 521-0378
;      TELEX:  LUKPAT WASHINGTON
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  648 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-468-718-2


  Query Match             13.3%;  Score 83;  DB 2;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 0.56;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
           : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  17
US-08-247-491A-2
; Sequence 2, Application US/08247491A
; Patent No. 5965400
;  GENERAL INFORMATION:
;    APPLICANT:  BRILES, David E.
;    APPLICANT:  YOTHER, Janet L.
;    TITLE OF INVENTION:  STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;      STREET:  745 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10151
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/247,491A
;      FILING DATE:  23-JUN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOWALSKI, Thomas J.
;      REGISTRATION NUMBER:  32,147
;      REFERENCE/DOCKET NUMBER:  454312-2041
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-588-0800
;      TELEFAX:  212-588-0500
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  648 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-247-491A-2


  Query Match             13.3%;  Score 83;  DB 2;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 0.56;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
           : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  18
US-08-446-201-3
; Sequence 3, Application US/08446201B
; Patent No. 6042838
; GENERAL INFORMATION:
;  APPLICANT: BRILES, David E.
;  APPLICANT: WU, Hong-Yin
;  TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS FOR MUCOSAL ADMINISTRATION OF
;  TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN A (PSPA)
;  FILE REFERENCE: 454312-2018
;  CURRENT APPLICATION NUMBER: US/08/446,201B
;  CURRENT FILING DATE: 1995-05-19
;  EARLIER APPLICATION NUMBER: 08/312,949
;  EARLIER FILING DATE: 1994-09-30
;  EARLIER APPLICATION NUMBER: 08/246,636
;  EARLIER FILING DATE: 1994-05-20
;  EARLIER APPLICATION NUMBER: 08/048,896
;  EARLIER FILING DATE: 1993-04-20
;  EARLIER APPLICATION NUMBER: 07/835,698
;  EARLIER FILING DATE: 1992-02-12
;  EARLIER APPLICATION NUMBER: 07/656,773
;  EARLIER FILING DATE: 1991-02-15
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 648
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
US-08-446-201-3


  Query Match             13.3%;  Score 83;  DB 3;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 0.56;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 413

Qy      66 DSGDYNMAKAKMKN 79
           : | :||   ::|
Db     414 YNTDGSMATGWLQN 427


RESULT  19
US-08-127-499A-23
; Sequence 23, Application US/08127499A
; Patent No. 5510264
;  GENERAL INFORMATION:
;    APPLICANT:  VAN ALSTYNE, Diane
;    APPLICANT:  SHARMA, Lawrence Rajendra
;    TITLE OF INVENTION:  ANTIBODIES WHICH BIND MENINGITIS RELATED
;    TITLE OF INVENTION:  HOMOLOGOUS ANTIGENIC SEQUENCES
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/127,499A
```

```
;      FILING DATE:  28-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  51916/102/INBI
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  695 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
US-08-127-499A-23


  Query Match             13.3%;  Score 83;  DB 1;  Length 695;
  Best Local Similarity   31.1%;  Pred. No. 0.61;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     396 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 455

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     456 YNTDGSMATGWLQN 469


RESULT  20
US-08-482-847-23
; Sequence 23, Application US/08482847
; Patent No. 5556757
;  GENERAL INFORMATION:
;    APPLICANT:  VAN ALSTYNE, Diane
;    APPLICANT:  SHARMA, Lawrence Rajendra
;    TITLE OF INVENTION:  PEPTIDES REPRESENTING EPITOPIC SITES FOR
;    TITLE OF INVENTION:  BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR
;    TITLE OF INVENTION:  CNS CARRIER, ANTIBODIES THERETO, AND USES THEREOF
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/482,847
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/127,499
;      FILING DATE:  28-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  51916/104/INBI
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  695 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
US-08-482-847-23


  Query Match             13.3%;  Score 83;  DB 1;  Length 695;
  Best Local Similarity   31.1%;  Pred. No. 0.61;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     396 EKPAEQPKPEKTDDQQAEEDYARRSEEEYNRLTQQQPPKAEKPAPAPKTGWKQENGMWYF 455

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     456 YNTDGSMATGWLQN 469


RESULT  21
US-08-363-215-2
; Sequence 2, Application US/08363215
; Patent No. 5807989
;  GENERAL INFORMATION:
;    APPLICANT:  Margolis, Benjamin L.
;    APPLICANT:  Schlessinger, Joseph
;    APPLICANT:  Yajnik, Vijay
;    TITLE OF INVENTION:  METHODS FOR TREATMENT OR
;    TITLE OF INVENTION:  DIAGNOSIS OF DISEASES OR
;    TITLE OF INVENTION:  DISORDERS ASSOCIATED
;    TITLE OF INVENTION:  WITH AN APB DOMAIN
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/363,215
;      FILING DATE:  December 23, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  212/208
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  473 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-363-215-2


  Query Match             13.0%;  Score 81.5;  DB 1;  Length 473;
  Best Local Similarity   26.2%;  Pred. No. 0.55;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   33;  Gaps    7;
```

```
Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|        | : :||  |  |       |: ||        :||
Db     229 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 288

Qy      47 KPGGSDFLMKRLQ-----KGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDHI 97
             ||    : |::        |:: ||    |    |: ||:   : :  |||    : |
Db     289 PVGGDPEVRKQMPPPPPCPGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS-- 343

Qy      98 PTPQDL 103
             |:||
Db     344 -APRDL 348


RESULT  22
US-08-807-342B-7
; Sequence 7, Application US/08807342B
; Patent No. 6077686
;  GENERAL INFORMATION:
;    APPLICANT:  Der, Channing
;    APPLICANT:  O'Bryan, John P.
;    APPLICANT:  Pawson, Anthony
;    TITLE OF INVENTION:  No. 6077686el SHC Proteins
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr., P.O. Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/807,342B
;      FILING DATE:  28-FEB-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/012,516
;      FILING DATE:  29-FEB-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  MTS2USA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9200
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  473 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-807-342B-7


  Query Match             13.0%;  Score 81.5;  DB 3;  Length 473;
  Best Local Similarity   26.2%;  Pred. No. 0.55;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   33;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|        | : :||  |  |       |: ||        :||
Db     229 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 288

Qy      47 KPGGSDFLMKRLQ-----KGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDHI 97
             ||    : |::        |:: ||    |    |: ||:   : :  |||    : |
Db     289 PVGGDPEVRKQMPPPPPCPGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS-- 343

Qy      98 PTPQDL 103
             |:||
Db     344 -APRDL 348


RESULT  23
US-08-664-962B-4
; Sequence 4, Application US/08664962B
; Patent No. 6218162
;  GENERAL INFORMATION:
;    APPLICANT:  Krystal, Gerald
;    TITLE OF INVENTION:  SH2-CONTAINING INOSITOL-PHOSPHATASE
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MERCHANT & GOULD
;      STREET:  3100 No. 6218162west Center, 90 South Seventh Street
;      CITY:  Minneapolis
;      STATE:  Minnesota
;      COUNTRY:  U.S.A.
;      ZIP:  55402-4131
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/664,962B
;      FILING DATE:  14-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mueller, Douglas P.
;      REGISTRATION NUMBER:  30,300
;      REFERENCE/DOCKET NUMBER:  M&G 7933.49-US-01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612-332-5300
;      TELEFAX:  612-332-9081
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 473 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-664-962B-4


  Query Match             13.0%;  Score 81.5;  DB 4;  Length 473;
  Best Local Similarity   26.2%;  Pred. No. 0.55;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   33;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|        | : :||  |  |       |: ||        :||
Db     229 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 288

Qy      47 KPGGSDFLMKRLQ-----KGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDHI 97
             ||    : |::        |:: ||    |    |: ||:   : :  |||    : |
Db     289 PVGGDPEVRKQMPPPPPCPGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS-- 343

Qy      98 PTPQDL 103
             |:||
Db     344 -APRDL 348


RESULT  24
US-09-311-743-4
; Sequence 4, Application US/09311743
; Patent No. 6238903
;  GENERAL INFORMATION:
;       APPLICANT: Krystal, Gerald
;       TITLE OF INVENTION: SH2-CONTAINING INOSITOL-PHOSPHATASE
;       NUMBER OF SEQUENCES: 24
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: BERESKIN & PARR
;            STREET: 40 King Street West
;            CITY: Toronto
```

```
;             STATE: Ontario
;             COUNTRY: Canada
;             ZIP: M5H 3Y2
;          COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;          CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/311,743
;             FILING DATE: 14-May-1999
;             CLASSIFICATION: <Unknown>
;          ATTORNEY/AGENT INFORMATION:
;             NAME: Gravelle, Micheline
;             REGISTRATION NUMBER: 40,261
;             REFERENCE/DOCKET NUMBER: 7771-32
;          TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 416-364-7311
;             TELEFAX: 416-361-1398
;    INFORMATION FOR SEQ ID NO: 4:
;          SEQUENCE CHARACTERISTICS:
;             LENGTH: 473 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-311-743-4


  Query Match             13.0%;  Score 81.5;  DB 4;  Length 473;
  Best Local Similarity   26.2%;  Pred. No. 0.55;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   33;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|       | : :||  |  |      |: ||       :||
Db     229 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 288

Qy      47 KPGGSDFLMKRLQ-----KGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDHI 97
             ||    : |::        |:: ||    |    |: ||:   : :  |||    : |
Db     289 PVGGDPEVRKQMPPPPPCPGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS-- 343

Qy      98 PTPQDL 103
             |:||
Db     344 -APRDL 348


RESULT  25
US-08-353-550-3
; Sequence 3, Application US/08353550
; Patent No. 5744313
;  GENERAL INFORMATION:
;    APPLICANT:  Williams, Lewis T.
;    APPLICANT:  Cavanaugh, William M.
;    TITLE OF INVENTION:  No. 5744313el Protein Domain Which Binds
;    TITLE OF INVENTION:  Tyrosine Phosphorylated Proteins
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/353,550
;      FILING DATE:  09-DEC-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dow, Karen B.
;      REGISTRATION NUMBER:  29,684
;      REFERENCE/DOCKET NUMBER:  02307K-057300
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-353-550-3


  Query Match             13.0%;  Score 81;  DB 1;  Length 423;
  Best Local Similarity   26.0%;  Pred. No. 0.54;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   34;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|       | : :||  |  |      |: ||       :||
Db     219 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 278

Qy      47 KPGGSDFLMKRL------QKGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDH 96
             ||    : |::        |:: ||    |    |: ||:   : :  |||    : |
Db     279 PVGGDPEVRKQMPPPPPCPAGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS- 334

Qy      97 IPTPQDL 103
             |:||
Db     335 --APRDL 339


RESULT  26
US-08-551-687-3
; Sequence 3, Application US/08551687
; Patent No. 5925547
;  GENERAL INFORMATION:
;    APPLICANT:  Williams, Lewis T.
;    APPLICANT:  Cavanaugh, William M.
;    TITLE OF INVENTION:  No. 5925547el Protein Domain Which Binds
;    TITLE OF INVENTION:  Tyrosine Phosphorylated Proteins
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/551,687
;      FILING DATE:  01-NOV-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/353,550
;      FILING DATE:  09-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy, Matthew B.
;      REGISTRATION NUMBER:  P39,787
;      REFERENCE/DOCKET NUMBER:  2307K-5731
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
```

```
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-551-687-3


  Query Match             13.0%;  Score 81;  DB 2;  Length 423;
  Best Local Similarity   26.0%;  Pred. No. 0.54;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   34;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|       | : :|| |  |      |: ||       :||
Db     219 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 278

Qy      47 KPGGSDFLMKRL------QKGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDH 96
             ||   : |::        |:: ||   |    |: ||:   : :  |||    : |
Db     279 PVGGDPEVRKQMPPPPPCPAGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS- 334

Qy      97 IPTPQDL 103
             |:||
Db     335 --APRDL 339


RESULT  27
US-08-353-550-8
; Sequence 8, Application US/08353550
; Patent No. 5744313
;  GENERAL INFORMATION:
;    APPLICANT:  Williams, Lewis T.
;    APPLICANT:  Cavanaugh, William M.
;    TITLE OF INVENTION:  No. 5744313el Protein Domain Which Binds
;    TITLE OF INVENTION:  Tyrosine Phosphorylated Proteins
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/353,550
;      FILING DATE:  09-DEC-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dow, Karen B.
;      REGISTRATION NUMBER:  29,684
;      REFERENCE/DOCKET NUMBER:  02307K-057300
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  474 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-353-550-8


  Query Match             13.0%;  Score 81;  DB 1;  Length 474;
  Best Local Similarity   26.0%;  Pred. No. 0.62;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   34;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|       | : :|| |  |      |: ||       :||
Db     229 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 288

Qy      47 KPGGSDFLMKRL------QKGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDH 96
             ||   : |::        |:: ||   |    |: ||:   : :  |||    : |
Db     289 PVGGDPEVRKQMPPPPPCPAGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS- 344

Qy      97 IPTPQDL 103
             |:||
Db     345 --APRDL 349


RESULT  28
US-08-551-687-8
; Sequence 8, Application US/08551687
; Patent No. 5925547
;  GENERAL INFORMATION:
;    APPLICANT:  Williams, Lewis T.
;    APPLICANT:  Cavanaugh, William M.
;    TITLE OF INVENTION:  No. 5925547el Protein Domain Which Binds
;    TITLE OF INVENTION:  Tyrosine Phosphorylated Proteins
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/551,687
;      FILING DATE:  01-NOV-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/353,550
;      FILING DATE:  09-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy, Matthew B.
;      REGISTRATION NUMBER:  P39,787
;      REFERENCE/DOCKET NUMBER:  2307K-5731
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  474 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-551-687-8


  Query Match             13.0%;  Score 81;  DB 2;  Length 474;
  Best Local Similarity   26.0%;  Pred. No. 0.62;
  Matches   33;  Conservative   18;  Mismatches   42;  Indels   34;  Gaps    7;

Qy       5 QEEENPAEE------TGEEKQ-----DTQEKEGILPERAEEAKLKAKYPS-------LGQ 46
           :||| | :        |:|       | : :|| |  |      |: ||       :||
Db     229 EEEEEPPDHQYYNDFPGKEPPLGGVVDMRLREGAAPGAARPTAPNAQTPSHLGATLPVGQ 288

Qy      47 KPGGSDFLMKRL------QKGQKYFDSGDY----NMAKAKMKNKQLPSAGPDKNLVTGDH 96
             ||   : |::        |:: ||   |    |: ||:   : :  |||    : |
Db     289 PVGGDPEVRKQMPPPPPCPAGRELFDDPSYVNVQNLDKAR---QAVGGAGPPNPAINGS- 344

Qy      97 IPTPQDL 103
             |:||
Db     345 --APRDL 349


RESULT  29
US-09-446-504-64
```

```
; Sequence 64, Application US/09446504
; Patent No. 6218150
; GENERAL INFORMATION:
;  APPLICANT:  UEMORI, Takashi
;  APPLICANT:  SATO, Yoshimi
;  APPLICANT:  FUJITA, Tomoko
;  APPLICANT:  MIYAKE, Kazue
;  APPLICANT:  MUKAI, Hiroyuki
;  APPLICANT:  ASADA, Kiyozo
;  APPLICANT:  KATO, Ikunoshin
;  TITLE OF INVENTION:  DNA POLYMERASE-RELATED FACTORS
;  FILE REFERENCE:  1422-408PCT
;  CURRENT APPLICATION NUMBER: US/09/446,504
;  CURRENT FILING DATE:  1999-12-23
;  PRIOR APPLICATION NUMBER:  PCT/JP98/02845
;  PRIOR FILING DATE:  1998-06-24
;  PRIOR APPLICATION NUMBER:  JP 9-187496
;  PRIOR FILING DATE:  1997-06-26
;  PRIOR APPLICATION NUMBER:  JP 9-320692
;  PRIOR FILING DATE:  1997-11-27
;  NUMBER OF SEQ ID NOS:  92
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 64
;   LENGTH: 479
;   TYPE: PRT
;   ORGANISM: Pyrococcus furiosus
US-09-446-504-64


  Query Match             12.7%;  Score 79.5;  DB 4;  Length 479;
  Best Local Similarity   38.5%;  Pred. No. 0.92;
  Matches   20;  Conservative    9;  Mismatches   20;  Indels    3;  Gaps    1;

Qy       5 QEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMK 56
           :|||   |:| |||:: :||     |  || | : : |  |::    ||| |
Db     431 EEEEEEEEKTEEEKEEIEEKP---EEEKEEEKKEKEKPKKGKQATLFDFLKK 479


RESULT  30
US-09-712-266-64
; Sequence 64, Application US/09712266
; Patent No. 6333158
; GENERAL INFORMATION:
;  APPLICANT: UEMORI, Takashi
;  APPLICANT:  SATO, Yoshimi
;  APPLICANT:  FUJITA, Tomoko
;  APPLICANT:  MIYAKE, Kazue
;  APPLICANT:  MUKAI, Hiroyuki
;  APPLICANT:  ASADA, Kiyozo
;  APPLICANT:  KATO, Ikunoshin
;  TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
;  FILE REFERENCE: 1422-408PCT
;  CURRENT APPLICATION NUMBER: US/09/712,266
;  CURRENT FILING DATE:  2000-11-15
;  PRIOR APPLICATION NUMBER: US 09/446,504
;  PRIOR FILING DATE: 1999-12-23
;  PRIOR APPLICATION NUMBER: PCT/JP98/02845
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: JP 9-187496
;  PRIOR FILING DATE: 1997-06-26
;  PRIOR APPLICATION NUMBER: JP 9-320692
;  PRIOR FILING DATE: 1997-11-27
;  NUMBER OF SEQ ID NOS: 92
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
;   LENGTH: 479
;   TYPE: PRT
;   ORGANISM: Pyrococcus furiosus
US-09-712-266-64


  Query Match             12.7%;  Score 79.5;  DB 4;  Length 479;
  Best Local Similarity   38.5%;  Pred. No. 0.92;
  Matches   20;  Conservative    9;  Mismatches   20;  Indels    3;  Gaps    1;

Qy       5 QEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMK 56
           :|||   |:| |||:: :||     |  || | : : |  |::    ||| |
Db     431 EEEEEEEEKTEEEKEEIEEKP---EEEKEEEKKEKEKPKKGKQATLFDFLKK 479


RESULT  31
US-08-741-134-6
; Sequence 6, Application US/08741134
; Patent No. 5861498
;  GENERAL INFORMATION:
;    APPLICANT:  Litwack, Gerald
;    APPLICANT:  Alnemri, Emad S.
;    APPLICANT:  Fernandes-Alnemri, Teresa
;    TITLE OF INVENTION:  IMMUNOPHILIN FKBP46 AND COMPOSITIONS FOR MAKING
;    TITLE OF INVENTION:  AND
;    TITLE OF INVENTION:  METHODS OF USING THE SAME
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5861498ris
;      STREET:  One Liberty Place - 46th floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Windows 3.11
;      SOFTWARE:  WordPerfect for Windows 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/741,134
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/007,163
;      FILING DATE:  01-NOV-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca, Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  TJU-2090
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 411 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-741-134-6


  Query Match             12.5%;  Score 78;  DB 2;  Length 411;
  Best Local Similarity   24.8%;  Pred. No. 1.1;
  Matches   32;  Conservative   21;  Mismatches   60;  Indels   16;  Gaps    3;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           : :: ||:|  ||  ||::: ::|| : ||  :  | | :     ::        |: :|
Db     219 VQEEDEEDNDGEEEQEEEEEEEEQKEEVKPEPKKSKKEKKRKHEEKEEE-------KKAKK 271

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHI----PTPQDLPQRKSSL-----V 111
           :|     |      | |:|:       |: |    |     |  | || |
Db     272 VKKVEFKKDLEEGPTKPKSKKEQDKHKPKSKVLEGGIVIEDRTIGDGPQAKRGARVGMRY 331

Qy     112 TSKLAGGQV 120
             ||  |:|
Db     332 IGKLKNGKV 340


RESULT  32
```

```
US-08-556-419-23
; Sequence 23, Application US/08556419C
; Patent No. 6093549
; GENERAL INFORMATION:
;  APPLICANT: Ross, Christopher
;  APPLICANT: Li, Xiao-Jiang
;  APPLICANT: Li, Shi-Hua
;  APPLICANT: Sharp, Alan
;  APPLICANT: Lanahan, Anthony
;  APPLICANT: Worley, Paul
;  APPLICANT: Snyder, Solomon
;  TITLE OF INVENTION: Huntingtin-associated protein
;  FILE REFERENCE: 01107.52271
;  CURRENT APPLICATION NUMBER: US/08/556,419C
;  CURRENT FILING DATE: 1995-11-09
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 23
;   LENGTH: 629
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-08-556-419-23
```

```
  Query Match             12.2%;  Score 76;  DB 3;  Length 629;
  Best Local Similarity   23.3%;  Pred. No. 3.1;
  Matches   28;  Conservative   22;  Mismatches   38;  Indels   32;  Gaps    5;

Qy       7 EENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSL---GQKPGGSD--FLMKRLQKG 61
           |: ||||| :  ::|:  | : ||  |  ::     :|   |  |   |  :::::|
Db     503 EDGPAEETEQASEETEAWEEVEPEVDEATRMNVVVSALEASGLGPSHLDMKYVLQQLSNW 562

Qy      62 QKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAGGQVE 121
           |          :| : | | :|   |                ||::::     : || ||
Db     563 QDAH-------SKRQQKQKVVPKDSP--------------APQQQTN-----MGGGIVE 595
```

```
RESULT  33
US-08-446-137B-2
; Sequence 2, Application US/08446137B
; Patent No. 6162903
;  GENERAL INFORMATION:
;    APPLICANT:  Trowern, Angus R.
;    APPLICANT:  Atkinson, Anthony
;    APPLICANT:  Murphy, Jonathan P.
;    APPLICANT:  Laurence, Oliver S.
;    APPLICANT:  Duggleby, Clive J.
;    TITLE OF INVENTION:  IMMUNOGLOBULIN BINDING PROTEINS DERIVED
;    TITLE OF INVENTION:  FROM L PROTEIN AND THEIR USES
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,137B
;      FILING DATE:  22-MAY-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
;      REFERENCE/DOCKET NUMBER:  100084.406
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
```

```
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1027 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-446-137B-2
```

```
  Query Match             12.1%;  Score 75.5;  DB 4;  Length 1027;
  Best Local Similarity   25.8%;  Pred. No. 6.7;
  Matches   23;  Conservative   19;  Mismatches   28;  Indels   19;  Gaps    3;

Qy       5 QEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKY 64
           :::| | :: ||:|:   :| |      |: | : | |   :||       :  : |:
Db     938 KKDEEPGKKPGEDKKPEDKKPG------EDKKPEDKKPGEDKKP-------EDKKPGKTD 984

Qy      65 FDSGDYNMAKAKMKNKQLPSAGPDKNLVT 93
             || :      | |  :|| || :  ::|
Db     985 KDSPN------KKKKAKLPKAGSEAEILT 1007
```

```
RESULT  34
5210183-2
;Patent No. 5210183
;    APPLICANT: LINDAHL, GUNNAR;FRITHZ, ELISABET;HEDEN, LARS-OLOF
;    TITLE OF INVENTION: PROTEIN ARP, WITH IMMUNOGLOBULIN A
;BINDING ACTIVITY, THE CORRESPONDING VECTORS AND HOSTS, REAGENT
;KIT AND PHARMACEUTICAL COMPOSITION
;    NUMBER OF SEQUENCES: 3
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/270,099
;      FILING DATE: 14-NOV-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 186,097
;      FILING DATE: 25-APR-1988
;SEQ ID NO:2:
;      LENGTH: 344
5210183-2
```

```
  Query Match             12.0%;  Score 75;  DB 6;  Length 344;
  Best Local Similarity   24.4%;  Pred. No. 1.9;
  Matches   30;  Conservative   24;  Mismatches   57;  Indels   12;  Gaps    4;

Qy       4 KQEEENPAEETGEEKQ-DTQEKEGILPE----RAEEAKLKAKYPSLGQKPGGSDFLMKRL 58
           |:| |  |::  ||||     ::|:  :    |  : |::|       |    : | | |
Db     185 KKELEAEAQKLKEEKQISDASRQGLSRDLEASRERKKKVEADLAEANSKLQALEKLNKEL 244

Qy      59 QKGQKYFDSGDYNM-----AKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQ--RKSSLV 111
           ::|:|  :     :      |:||   :||      :   : |:  | :  ||  |  | :
Db     245 EEGKKLSEKEKAELQARLEAEAKALKEQLAKQAEELAKLKGNQTPNAKVAPQANRSRSAM 304

Qy     112 TSK 114
           | :
Db     305 TQQ 307
```

```
RESULT  35
5210183-3
;Patent No. 5210183
;    APPLICANT: LINDAHL, GUNNAR;FRITHZ, ELISABET;HEDEN, LARS-OLOF
;    TITLE OF INVENTION: PROTEIN ARP, WITH IMMUNOGLOBULIN A
;BINDING ACTIVITY, THE CORRESPONDING VECTORS AND HOSTS, REAGENT
;KIT AND PHARMACEUTICAL COMPOSITION
;    NUMBER OF SEQUENCES: 3
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/270,099
;      FILING DATE: 14-NOV-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 186,097
;      FILING DATE: 25-APR-1988
;SEQ ID NO:3:
```

```
;      LENGTH: 683
5210183-3


  Query Match             12.0%;  Score 75;  DB 6;  Length 683;
  Best Local Similarity   24.4%;  Pred. No. 4.5;
  Matches   30;  Conservative   24;  Mismatches   57;  Indels   12;  Gaps    4;

Qy       4 KQEEENPAEETGEEKQ-DTQEKEGILPE----RAEEAKLKAKYPSLGQKPGGSDFLMKRL 58
           |:| |   |::  ||||     ::|:  :    |  : |::|       |    : | | |
Db     185 KKELEAEAQKLKEEKQISDASRQGLSRDLEASRERKKKVEADLAEANSKLQALEKLNKEL 244

Qy      59 QKGQKYFDSGDYNM-----AKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQ--RKSSLV 111
           ::|:|  :     :     |:||   :||     :   : |:  |  :  ||  |  | :
Db     245 EEGKKLSEKEKAELQARLEAEAKALKEQLAKQAEELAKLKGNQTPNAKVAPQANRSRSAM 304

Qy     112 TSK 114
           | :
Db     305 TQQ 307


RESULT  36
US-08-938-291A-9
; Sequence 9, Application US/08938291A
; Patent No. 6117673
;  GENERAL INFORMATION:
;    APPLICANT:  Lev, Sima
;    APPLICANT:  Plowman, Gregory D.
;    APPLICANT:  Schlessinger, Joseph
;    TITLE OF INVENTION:  RDGB PROTEINS AND RELATED
;    TITLE OF INVENTION:  PRODUCTS AND METHODS
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSeq
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/938,291A
;      FILING DATE:  September 26, 1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/027,337
;      FILING DATE:  October 11, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  228/172
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1250 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-938-291A-9


  Query Match             11.8%;  Score 73.5;  DB 3;  Length 1250;
  Best Local Similarity   23.8%;  Pred. No. 14;
  Matches   34;  Conservative   18;  Mismatches   50;  Indels   41;  Gaps    6;

Qy       3 QKQEEENPAEETGEEKQDTQEKE-------------GILPERAEEAKL----KAKYPSL 44
           |    |  ::| |||::|  | |              |   :|:    :    | |: |
Db     311 QPPSAEASSDEEGEEEEDDDEDENDAIGTGVDLSANQGGSAQRSRSQSIQMAQKGKFGSK 370

Qy      45 G--QKPGGS---------DFLMKRLQ------KGQKYFDSGDYNMAKAKMKNKQLPSAGP 87
           |    | ||         :: |:||:        :::||  | |   :  |   |   |
Db     371 GALHSPVGSAHSFDLQVANWRMERLEVDSKSNSDEEFFDCLDTNETNSLAKWSSLELLG- 429

Qy      88 DKNLVTGDHIPTPQDLPQRKSSL 110
             ||  | |   |   :|:
Db     430 -----EGDDSPPPHGGPSSAASV 447


RESULT  37
US-09-136-574A-43
; Sequence 43, Application US/09136574A
; Patent No. 6294366
;   GENERAL INFORMATION:
;        APPLICANT: Farrington, Graham K.
;                   Anderson, Paige
;                   Gibbs, Moreland
;                   Bergquist, Peter
;                   Daniels, Roy
;                   Morgan, Hugh W.
;                   Williams, Diane P.
;        TITLE OF INVENTION: Compositions and Methods for
;                            Treating Cellulose Containing Fabrics Using Truncated
;                            Cellulase Enzyme Compositions
;        NUMBER OF SEQUENCES: 49
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Howson and Howson
;             STREET: Spring House Corporate Center, P.O. Box 457
;             CITY: Spring House
;             STATE: PA
;             COUNTRY: USA
;             ZIP: 19477
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/136,574A
;             FILING DATE: 19-Aug-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 08/932,571
;             FILING DATE: September 19, 1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Bak, Mary E.
;             REGISTRATION NUMBER: 31,215
;             REFERENCE/DOCKET NUMBER: 1997US001/CIP
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 215-540-9200
;             TELEFAX: 215-540-5818
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 43:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1426 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: No. 6294366e
;        SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-09-136-574A-43


  Query Match             11.6%;  Score 72.5;  DB 4;  Length 1426;
  Best Local Similarity   23.4%;  Pred. No. 22;
  Matches   30;  Conservative   16;  Mismatches   39;  Indels   43;  Gaps    6;

Qy      15 GEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMK----------------- 56
```

```
          ||  |       | |     | :  |:  ||   ||:|: ::
Db    827 GEATQSVSVTSSINP-----AYIDVKFVKLGANAGGADYYVEIGFKSGAGVLAAGQSTKE 881

Qy     57 ---RLQKGQ-KYFDSGDYNM--AKAKMKNKQL------------PSAG-PDKNLVTGDH 96
           :|||   |  | ||::  | : ::|:::             || |     :||
Db    882 IRLSIQKGSGSYNQSNDYSIRSANSYIENEKVTGYIDGAIVWGREPSRGTKPAGVVTPTP 941

Qy     97 IPTPQDLP 104
          |||   |
Db    942 APTPTSTP 949


RESULT  38
US-08-285-440-4
; Sequence 4, Application US/08285440
; Patent No. 5532337
;  GENERAL INFORMATION:
;    APPLICANT:  Ken'ichiro HAYASHI et al.
;    TITLE OF INVENTION:  POLYPEPTIDE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 500 kb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  DisplayWrite
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/285,440
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/858,947
;      FILING DATE:  March 27, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  312 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;      ORGANISM:
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      HAPLOTYPE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:
;      ORGANELLE:
;    IMMEDIATE SOURCE:
;      LIBRARY:
;      CLONE:
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:
;      MAP POSITION:
;      UNITS:
;    FEATURE:
;      NAME/KEY:
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
;    PUBLICATION INFORMATION:
;      AUTHORS:
;      TITLE:
;      JOURNAL:
;      VOLUME:
;      ISSUE:
;      PAGES:
;      DATE:
;      DOCUMENT NUMBER:
;      FILING DATE:
;      PUBLICATION DATE:
;      RELEVANT RESIDUES IN SEQ ID NO:
US-08-285-440-4


  Query Match             11.5%;  Score 72;  DB 1;  Length 312;
  Best Local Similarity   24.5%;  Pred. No. 3.5;
  Matches   39;  Conservative   20;  Mismatches   44;  Indels   56;  Gaps    7;

Qy      3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLG----QKP-------GGS 51
          |::::|    :  ||::  : || |   ||| |: : | |  |    :||       | |
Db    103 QRRKQEEADRKLREEEEKRRLKEEIERRRAEAAEKRQKMPEDGLSDDKKPFKCFTPKGSS 162

Qy     52 -------DFLMKRLQKGQKYFDSGDYNMAKAKMKNK------------QLPSAGPDKNL 91
                :|| | :||      ||  :  :| : :|               || | |
Db    163 LKIEERAEFLNKSVQK-----SSGVKSTHQAAIVSKIDSRLEQYTSAIEGTKSAKPTK-- 215

Qy     92 VTGDHIPTPQDLP-----------QRKSSLVTSKLAGG 118
              |  |||             | :: :|  | |
Db    216 ------PAASDLPVPAEGVRNIKSMWEKGNVFSSPTAAG 248


RESULT  39
US-08-630-349-4
; Sequence 4, Application US/08630349
; Patent No. 5739008
;  GENERAL INFORMATION:
;    APPLICANT:  Ken'ichiro HAYASHI et al.
;    TITLE OF INVENTION:  POLYPEPTIDE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 144 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/630,349
;      FILING DATE:  April 10, 1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/285,440
;      FILING DATE:  August 4, 1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/858,947
;      FILING DATE:  March 27, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  312 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-630-349-4


  Query Match             11.5%;  Score 72;  DB 1;  Length 312;
  Best Local Similarity   24.5%;  Pred. No. 3.5;
  Matches   39;  Conservative   20;  Mismatches   44;  Indels   56;  Gaps    7;

Qy      3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLG----QKP-------GGS 51
          |::::|    :  ||::  : || |   ||| |: : | |  |    :||       | |
Db    103 QRRKQEEADRKLREEEEKRRLKEEIERRRAEAAEKRQKMPEDGLSDDKKPFKCFTPKGSS 162

Qy     52 -------DFLMKRLQKGQKYFDSGDYNMAKAKMKNK------------QLPSAGPDKNL 91
                :|| | :||      ||  :  :| : :|               || | |
Db    163 LKIEERAEFLNKSVQK-----SSGVKSTHQAAIVSKIDSRLEQYTSAIEGTKSAKPTK-- 215

Qy     92 VTGDHIPTPQDLP-----------QRKSSLVTSKLAGG 118
            |   |||               | :: :|  | |
Db    216 ------PAASDLPVPAEGVRNIKSMWEKGNVFSSPTAAG 248


RESULT  40
US-08-285-440-5
; Sequence 5, Application US/08285440
; Patent No. 5532337
;  GENERAL INFORMATION:
;    APPLICANT:  Ken'ichiro HAYASHI et al.
;    TITLE OF INVENTION:  POLYPEPTIDE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 500 kb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  DisplayWrite
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/285,440
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/858,947
;      FILING DATE:  March 27, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  532 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;      ORGANISM:
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      HAPLOTYPE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:
;      ORGANELLE:
;    IMMEDIATE SOURCE:
;      LIBRARY:
;      CLONE:
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:
;      MAP POSITION:
;      UNITS:
;    FEATURE:
;      NAME/KEY:
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
;    PUBLICATION INFORMATION:
;      AUTHORS:
;      TITLE:
;      JOURNAL:
;      VOLUME:
;      ISSUE:
;      PAGES:
;      DATE:
;      DOCUMENT NUMBER:
;      FILING DATE:
;      PUBLICATION DATE:
;      RELEVANT RESIDUES IN SEQ ID NO:
US-08-285-440-5


  Query Match             11.5%;  Score 72;  DB 1;  Length 532;
  Best Local Similarity   24.5%;  Pred. No. 6.9;
  Matches   39;  Conservative   20;  Mismatches   44;  Indels   56;  Gaps    7;

Qy      3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLG----QKP-------GGS 51
          |::::|    :  ||::  : || |   ||| |: : | |  |    :||       | |
Db    323 QRRKQEEADRKLREEEEKRRLKEEIERRRAEAAEKRQKMPEDGLSDDKKPFKCFTPKGSS 382

Qy     52 -------DFLMKRLQKGQKYFDSGDYNMAKAKMKNK------------QLPSAGPDKNL 91
                :|| | :||      ||  :  :| : :|               || | |
Db    383 LKIEERAEFLNKSVQK-----SSGVKSTHQAAIVSKIDSRLEQYTSAIEGTKSAKPTK-- 435

Qy     92 VTGDHIPTPQDLP-----------QRKSSLVTSKLAGG 118
            |   |||               | :: :|  | |
Db    436 ------PAASDLPVPAEGVRNIKSMWEKGNVFSSPTAAG 468


Search completed: September  7, 2002, 10:43:51
Job time: 247 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:45:21 ; Search time 51.07 Seconds
(without alignments)
263.167 Million cell updates/sec

Title: US-09-824-178-4
Perfect score: 625
Sequence: 1 MSQKQEEENPAEETGEEKQD.....DLPQRKSSLVTSKLAGGQVE 121

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	625	100.0	121	19	AAW61552	Human endosulfine
2	625	100.0	121	22	AAB93514	Human protein sequ
3	607	97.1	122	21	AAG01489	Human secreted pro
4	605	96.8	117	19	AAW61551	Human endosulfine
5	492	78.7	150	21	AAB54082	Human pancreatic c
6	471	75.4	118	19	AAW60214	Human phosphoprote
7	427.5	68.4	125	22	AAO13512	Human polypeptide
8	402	64.3	96	19	AAW62655	Homo sapiens C4-2
9	296.5	47.4	82	21	AAG00770	Human secreted pro
10	296.5	47.4	82	21	AAG00772	Human secreted pro
11	288	46.1	101	21	AAG00771	Human secreted pro
12	283	45.3	66	21	AAG00769	Human secreted pro
13	226	36.2	119	22	ABB71486	Drosophila melanog
14	108	17.3	605	21	AAY57950	Human transmembran
15	104.5	16.7	200	18	AAW14545	Streptococcus pneu
16	102	16.3	605	20	AAY31741	Human podocalyxin-
17	102	16.3	605	22	AAU12251	Human PRO3742 poly
18	102	16.3	605	22	AAE03767	Human gene 4 encod
19	97.5	15.6	211	18	AAW14547	Streptococcus pneu
20	94.5	15.1	211	18	AAW14542	Streptococcus pneu
21	90.5	14.5	360	22	AAM93868	Human polypeptide,
22	89	14.2	193	22	ABG11252	Novel human diagno
23	87.5	14.0	1751	22	ABB69600	Drosophila melanog
24	85	13.6	615	22	ABG25869	Novel human diagno
25	85	13.6	659	22	ABG09291	Novel human diagno
26	84.5	13.5	131	22	ABG04385	Novel human diagno
27	84	13.4	147	22	ABG11269	Novel human diagno
28	84	13.4	969	22	ABB68123	Drosophila melanog
29	84	13.4	2286	22	AAB65635	Novel protein kina
30	83.5	13.4	281	22	ABB71057	Drosophila melanog
31	83.5	13.4	374	21	AAG16884	Arabidopsis thalia
32	83	13.3	619	15	AAR63437	Pneumococcal surfa
33	83	13.3	619	17	AAR87598	Pneumococcal surfa
34	83	13.3	619	17	AAR86911	Pneumococcal surfa
35	83	13.3	619	20	AAY41838	Streptococcus pneu
36	83	13.3	648	19	AAW70336	Pneumococcal surfa
37	83	13.3	648	19	AAW62274	Streptococcus pneu
38	83	13.3	648	20	AAY41837	Streptococcus pneu
39	83	13.3	648	20	AAW87879	A pneumococcal sur
40	83	13.3	653	20	AAW92456	S. pneumoniae trun
41	83	13.3	670	21	AAG51795	Arabidopsis thalia
42	83	13.3	684	16	AAR73912	Streptococcus pneu
43	83	13.3	688	21	AAG51794	Arabidopsis thalia
44	83	13.3	718	21	AAG42569	Arabidopsis thalia
45	83	13.3	722	21	AAG51793	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAW61552
ID AAW61552 standard; Protein; 121 AA.
XX
AC AAW61552;
XX
DT 19-OCT-1998 (first entry)
XX
DE Human endosulfine B protein.
XX
KW Endosulfine; central nervous system; pancreatic beta-cell; pain;
KW insulin; cytoprotective agent; diabetes; epilepsy; depression; ischemia.
XX
OS Homo sapiens.
XX
PN WO9830692-A2.
XX
PD 16-JUL-1998.
XX
PF 07-JAN-1998; 98WO-US00137.
XX
PR 07-JAN-1997; 97US-0779775.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Anderson KL, Roch J, Scott VE, Sullivan JP;
XX
DR WPI; 1998-399140/34.
DR N-PSDB; AAV45211.
XX
PT Novel human endosulfine A and B - for screening for compounds
PT modulating endosulfine receptor activity, useful as cytoprotective
PT agent(s)
XX

PS Claim 9; Page 58; 76pp; English.
XX
CC The human endosulfine A and B genes encode proteins related to other
CC members of the endosulfine family which have been shown to bind receptors
CC from both the central nervous system and pancreatic beta-cells and to
CC induce insulin secretion from a rodent beta-cell tumour line. The
CC protein can be used to screen for compounds that modulate endosulfine
CC receptor activity, especially as a cytoprotective agent.
CC Modulators of endosulfine activity may represent a treatment of
CC diabetes, epilepsy, pain, depression and ischemia.
XX
SQ Sequence 121 AA;

Query Match 100.0%; Score 625; DB 19; Length 121;
Best Local Similarity 100.0%; Pred. No. 7.6e-60;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
Db 1 msqkqeeenpaeetgeekqdtqekegilperaeeaklkakypslgqkpggsdflmkrlqk 60

Qy 61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAGGQV 120
Db 61 gqkyfdsgdynmakakmknkqlpsagpdknlvtgdhiptpqdlpqrksslvtsklaggqv 120

Qy 121 E 121
Db 121 e 121

RESULT 2
AAB93514
ID AAB93514 standard; Protein; 121 AA.
XX
AC AAB93514;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12846.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 12846; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 121 AA;

Query Match 100.0%; Score 625; DB 22; Length 121;
Best Local Similarity 100.0%; Pred. No. 7.6e-60;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
Db 1 msqkqeeenpaeetgeekqdtqekegilperaeeaklkakypslgqkpggsdflmkrlqk 60

Qy 61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAGGQV 120
Db 61 gqkyfdsgdynmakakmknkqlpsagpdknlvtgdhiptpqdlpqrksslvtsklaggqv 120

Qy 121 E 121
Db 121 e 121

RESULT 3
AAG01489
ID AAG01489 standard; Protein; 122 AA.
XX
AC AAG01489;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 5570.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR N-PSDB; AAC01495.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 13; SEQ ID 5570; 71pp + CD-ROM; English.
XX
CC The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
SQ Sequence 122 AA;

Query Match 97.1%; Score 607; DB 21; Length 122;
Best Local Similarity 97.5%; Pred. No. 6.8e-58;
Matches 117; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 msqkqeeenpaeetgeekqdtqekegilperaeeaklkakypslgqkpggsdflmkrlqk 60

Qy 61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAGGQV 120
|||||||||||||||||||||||||||||||||| |||||||||||| |||||||:
Db 61 gqkyfdsgdynmakakmknkqlpsagpdknlvtgdhipxpqdlpqrksslvxsklaggql 120

RESULT 4
AAW61551
ID AAW61551 standard; Protein; 117 AA.
XX
AC AAW61551;
XX
DT 19-OCT-1998 (first entry)
XX
DE Human endosulfine A protein.
XX
KW Endosulfine; central nervous system; pancreatic beta-cell; pain;
KW insulin; cytoprotective agent; diabetes; epilepsy; depression; ischemia.
XX
OS Homo sapiens.
XX
PN WO9830692-A2.
XX
PD 16-JUL-1998.
XX
PF 07-JAN-1998; 98WO-US00137.
XX
PR 07-JAN-1997; 97US-0779775.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Anderson KL, Roch J, Scott VE, Sullivan JP;
XX
DR WPI; 1998-399140/34.
DR N-PSDB; AAV45210.
XX
PT Novel human endosulfine A and B - for screening for compounds
PT modulating endosulfine receptor activity, useful as cytoprotective
PT agent(s)
XX
PS Claim 9; Page 58-59; 76pp; English.
XX
CC The human endosulfine A and B genes encode proteins related to other
CC members of the endosulfine family which have been shown to bind receptors
CC from both the central nervous system and pancreatic beta-cells and to
CC induce insulin secretion from a rodent beta-cell tumour line. The
CC protein can be used to screen for compounds that modulate endosulfine
CC receptor activity, especially as a cytoprotective agent.
CC Modulators of endosulfine activity may represent a treatment of
CC diabetes, epilepsy, pain, depression and ischemia.
XX
SQ Sequence 117 AA;

Query Match 96.8%; Score 605; DB 19; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.1e-57;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 msqkqeeenpaeetgeekqdtqekegilperaeeaklkakypslgqkpggsdflmkrlqk 60

Qy 61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 gqkyfdsgdynmakakmknkqlpsagpdknlvtgdhiptpqdlpqrksslvtsklag 117

RESULT 5
AAB54082
ID AAB54082 standard; Protein; 150 AA.
XX
AC AAB54082;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human pancreatic cancer antigen protein sequence SEQ ID NO:534.
XX
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative.
XX
OS Homo sapiens.
XX
PN WO200055320-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05989.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-579444/54.
DR N-PSDB; AAC98847.
XX
PT New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX
PS Claim 11; Page 972-973; 1379pp; English.
XX
CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to

```
CC   detect, treat or prevent pancreatic disorders, especially cancer.
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   150 AA;


  Query Match             78.7%;  Score 492;  DB 21;  Length 150;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-45;
  Matches   95;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      23 EKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQL 82
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56 ekegilperaeeaklkakypslgqkpggsdflmkrlqkgqkyfdsgdynmakakmknkql 115

Qy      83 PSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           ||||||||||||||||||||||||||||||||||
Db     116 psagpdknlvtgdhiptpqdlpqrksslvtsklag 150


RESULT   6
AAW60214
ID   AAW60214 standard; Protein; 118 AA.
XX
AC   AAW60214;
XX
DT   28-SEP-1998  (first entry)
XX
DE   Human phosphoprotein hPSHP.
XX
KW   hPSHP; phosphoprotein; neurotransmitter; signal transduction;
KW   human; Parkinson's disease; Huntington's diseases; diagnosis;
KW   drug screening.
XX
OS   Homo sapiens.
XX
PN   WO9818922-A1.
XX
PD   07-MAY-1998.
XX
PF   14-OCT-1997;   97WO-US18489.
XX
PR   28-OCT-1996;   96US-0739484.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Hillman JL;
XX
DR   WPI; 1998-272217/24.
DR   N-PSDB; AAV30285.
XX
PT   Human phosphoprotein polypeptide, hPSHP - useful for treating
PT   degenerative neuronal diseases e.g. Parkinson's and Huntington's,
PT   diagnosing hPSHP related diseases and drug screening
XX
PS   Claim 1; Page 48-49; 64pp; English.
XX
CC   This polypeptide comprises hPSHP, a novel human phosphoprotein that
CC   is believed to function in the signal transduction pathway of
CC   neurotransmitters in brain tissue and in cell development processes
CC   in peripheral organs, based on its chemical and structural homology
CC   with known phosphoproteins GI 162690 (71% identity) and GI 741603
CC   83% identity).  Its amino acid sequence was deduced from an
CC   isolated genomic DNA fragment (see V302850.  The invention also
CC   provides genetically engineered expression vectors, and host cells
CC   comprising nucleic acid sequences encoding hPSHP used in a claimed
CC   method for producing hPSHP, as well as probes and primers,
CC   antibodies, agonists and antagonists of hPSHP.  hPSHP (or its
CC   fragments) is useful therapeutically to regenerate and enhance
CC   survival of nerve cells to halt degenerative processes in brain
CC   diseases such as Parkinson's and Huntington's disease.  It can be
CC   used to induce antibodies to hPSHP, useful to diagnose conditions
CC   or diseases characterised by hPSHP expression, in assays to monitor
CC   patients receiving treatment, as agonists, antagonists or targeting
CC   molecules for pharmaceutical delivery, or to purify naturally
CC   occurring or recombinant hPSHP.  The polypeptide or fragments are
CC   also useful in conventional drug screening.  hPSHP agonists are
CC   useful e.g. for stimulating residual hPSHP in treatment of
CC   degenerative neuronal diseases, whilst antagonists are useful for
CC   inhibiting hPSHP, e.g. to regulate cell growth or suppress abnormal
CC   signal transduction in diseased tissue.
XX
SQ   Sequence   118 AA;


  Query Match             75.4%;  Score 471;  DB 19;  Length 118;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-43;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      31 RAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKN 90
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 raeeaklkakypslgqkpggsdflmkrlqkgqkyfdsgdynmakakmknkqlpsagpdkn 87

Qy      91 LVTGDHIPTPQDLPQRKSSLVTSKLAGGQVE 121
           |||||||||||||||||||||||||||||||
Db      88 lvtgdhiptpqdlpqrksslvtsklaggqve 118


RESULT   7
AAO13512
ID   AAO13512 standard; Protein; 125 AA.
XX
AC   AAO13512;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 27404.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI93443.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 20; SEQ ID NO 27404; 1399pp + Sequence Listing; English.
XX
```

```
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   125 AA;


  Query Match             68.4%;  Score 427.5;  DB 22;  Length 125;
  Best Local Similarity   76.4%;  Pred. No. 1.8e-38;
  Matches   84;  Conservative    9;  Mismatches   16;  Indels    1;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDS 67
           | |   : ||::: ::|    ||:|||||||:|| ||||||||||| ||||||||||
Db      17 evpeaasaeeqkemedkv-tspekaeeaklkaryphlgqkpggsdflrkrlqkgqkyfds 75

Qy      68 GDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           ||||||||||||||||:| |||  |||||||||||||||| ||| |||||
Db      76 gdynmakakmknkqlptaapdktevtgdhiptpqdlpqrkpslvasklag 125


RESULT   8
AAW62655
ID   AAW62655 standard; Protein; 96 AA.
XX
AC   AAW62655;
XX
DT   27-OCT-1998  (first entry)
XX
DE   Homo sapiens C4-2 tumour suppressor protein.
XX
KW   C4-2; tumour suppressor gene; cancer; treatment; glioblastoma;
KW   cell proliferative disorders; brain tumour.
XX
OS   Homo sapiens.
XX
PN   WO9820021-A1.
XX
PD   14-MAY-1998.
XX
PF   07-NOV-1997;   97WO-US20408.
XX
PR   08-NOV-1996;   96US-0744905.
XX
PA   (NWBI-) NORTHWEST BIOTHERAPEUTICS LLC.
XX
PI   Boynton AL,  Murphy GP,  Sehgal A;
XX
DR   WPI; 1998-286852/25.
DR   N-PSDB; AAV38840.
XX
PT   Tumour suppressor, C4-2 - useful for treatment of proliferative
PT   diseases e.g. cancer
XX
PS   Example; Fig 8; 112pp; English.
XX
CC   The sequence is that encoded by the C4-2 tumour suppressor gene.
CC   It can be used in the treatment of cell proliferative diseases,
CC   e.g. cancer, in particular glioblastomas and other malignant
CC   brain tumours.
XX
SQ   Sequence   96 AA;


  Query Match             64.3%;  Score 402;  DB 19;  Length 96;
  Best Local Similarity   86.5%;  Pred. No. 7.3e-36;
  Matches   77;  Conservative    4;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      29 PERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPD 88
           ||:|||||||:|| ||||||||:|| ||||||||||||||||||||||||||:| ||
Db       8 pekaeeaklkaryphlgqkpggsnflrkrlqkgqkyfdsgdynmakakmknkqlptaapd 67

Qy      89 KNLVTGDHIPTPQDLPQRKSSLVTSKLAG 117
           |  |||||||||||||| | ||| |||||
Db      68 ktevtgdhiptpqdlpqlkpslvasklag 96


RESULT   9
AAG00770
ID   AAG00770 standard; Protein; 82 AA.
XX
AC   AAG00770;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 4851.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC00776.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 4851; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   82 AA;


  Query Match             47.4%;  Score 296.5;  DB 21;  Length 82;
  Best Local Similarity   72.5%;  Pred. No. 1.5e-24;
  Matches   58;  Conservative    9;  Mismatches   12;  Indels    1;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDS 67
           | |   : ||::: ::|    ||:|||||||:|| |||||||||| ||||||||||
```

```
Db      4 evpeaasaeeqkemedkv-tspekaeeaklkaryphlgqkpggsdflrkrlqkgqkyfds 62

Qy     68 GDYNMAKAKMKNKQLPSAGP 87
          |||||||||||||||:| |
Db     63 gdynmakakmknkqlptaap 82


RESULT  10
AAG00772
ID   AAG00772 standard; Protein; 82 AA.
XX
AC   AAG00772;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 4853.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC00778.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 4853; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   82 AA;


  Query Match             47.4%;  Score 296.5;  DB 21;  Length 82;
  Best Local Similarity   72.5%;  Pred. No. 1.5e-24;
  Matches   58;  Conservative    9;  Mismatches   12;  Indels    1;  Gaps    1;

Qy      8 ENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDS 67
          | |   : ||::: ::|    ||:|||||||:|| |||||||||| ||||||||||
Db      4 evpeaasaeeqkemedkv-tspekaeeaklkaryphlgqkpggsdflrkrlqkgqkyfds 62

Qy     68 GDYNMAKAKMKNKQLPSAGP 87
          |||||||||||||||:| |
Db     63 gdynmakakmknkqlptaap 82


RESULT  11
AAG00771
ID   AAG00771 standard; Protein; 101 AA.
XX
AC   AAG00771;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 4852.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC00777.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 4852; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   101 AA;


  Query Match             46.1%;  Score 288;  DB 21;  Length 101;
  Best Local Similarity   60.4%;  Pred. No. 1.6e-23;
  Matches   61;  Conservative    6;  Mismatches   20;  Indels   14;  Gaps    1;

Qy      1 MSQKQEEENPAEETGEEKQ----------DTQEKEGILPERAEEAKLKAKYPSLGQ 46
          || :  |   |||   |               :  | :   ||:|||||||:|| |||
Db      1 msaevpeaasaeeqksehnmlpwslqpsipnsleemedkvtspekaeeaklkaryphlgq 60

Qy     47 KPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGP 87
          |||||||| |||||||||||||||||||||||||||:| |
Db     61 kpggsdflrkrlqkgqkyfdsgdynmakakmknkqlptaap 101


RESULT  12
AAG00769
ID   AAG00769 standard; Protein; 66 AA.
XX
AC   AAG00769;
XX
DT   06-OCT-2000  (first entry)
```

```
XX
DE   Human secreted protein, SEQ ID NO: 4850.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC00775.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 4850; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   66 AA;


  Query Match             45.3%;  Score 283;  DB 21;  Length 66;
  Best Local Similarity   89.8%;  Pred. No. 3.3e-23;
  Matches   53;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy     29 PERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGP 87
           ||:|||||||:|| |||||||||| |||||||||||||||||||||||||||:| |
Db      8 pekaeeaklkaryphlgqkpggsdflrkrlqkgqkyfdsgdynmakakmknkqlptaap 66


RESULT  13
ABB71486
ID   ABB71486 standard; Protein; 119 AA.
XX
AC   ABB71486;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 41250.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL15589.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 41250; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   119 AA;


  Query Match             36.2%;  Score 226;  DB 22;  Length 119;
  Best Local Similarity   43.7%;  Pred. No. 1e-16;
  Matches   52;  Conservative   17;  Mismatches   32;  Indels   18;  Gaps    3;

Qy      1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGG-SDFLMKRLQ 59
          ||  :|  |    | :: : |::      |: || |||:|||| : ||| | || ||||
Db      1 mssaeensnspattpqdtetteqanltdlekieeeklkskypsgmrvpgghsaflqkrlq 60

Qy     60 KGQKYFDSGDYNMAKAK-------MKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLV 111
          ||||:|||||| ||| |         ||          : ||: ||||: :| ||:|::
Db     61 kgqkffdsgdygmakqkgggvkqvfank----------vttgeaiptpetvparktsii 109


RESULT  14
AAY57950
ID   AAY57950 standard; Protein; 605 AA.
XX
AC   AAY57950;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Human transmembrane protein HTMPN-74.
XX
KW   Human; transmembrane protein; HTMPN; diagnosis; immunospecific;
KW   antiproliferative; neuroprotective; immune disorder;
KW   reproductive disorder; smooth muscle disorder; neurological disorder;
KW   gastrointestinal disorder; developmental disorder;
KW   cell proliferative disorder.
XX
OS   Homo sapiens.
XX
PN   WO9961471-A2.
XX
PD   02-DEC-1999.
XX
PF   28-MAY-1999;   99WO-US11904.
XX
PR   29-MAY-1998;   98US-0087260.
PR   02-JUL-1998;   98US-0091674.
```

```
PR   02-OCT-1998;   98US-0102954.
PR   24-NOV-1998;   98US-0109869.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Tang YT,  Lal P,  Hillman JL,  Yue H,  Guegler KJ,  Corley NC;
PI   Bandman O,  Patterson C,  Gorgone GA,  Kaser MR,  Baughn MR;
PI   Au-Young J;
XX
DR   WPI; 2000-072605/06.
DR   N-PSDB; AAZ56771.
XX
PT   Proteins, polynucleotides, vectors, host cells and antibodies used to
PT   diagnose, treat or prevent immune, reproductive, smooth muscle,
PT   neurological, gastrointestinal, developmental and cell proliferative
PT   disorders -
XX
PS   Claim 1; Page 172-174; 229pp; English.
XX
CC   AAZ56698 to AAZ56776 encode AAY57877 to AAY57955 which represent human
CC   transmembrane proteins designated HTMPN-1 to HTMPN-79, respectively.
CC   The transmembrane protein have immunospecific, antiproliferative and
CC   neuroprotective activities. The human transmembrane proteins,
CC   polynucleotides encoding them and other compositions and methods from
CC   the present invention, can be used for the diagnosis, treatment or
CC   prevention of immune, reproductive, smooth muscle, neurological,
CC   gastrointestinal, developmental and cell proliferative disorders. The
CC   HTMPN's can be used to treat or prevent disorders associated with a
CC   decreased expression or activity of HTMPN.
XX
SQ   Sequence   605 AA;
```

```
  Query Match             17.3%;  Score 108;  DB 21;  Length 605;
  Best Local Similarity   29.7%;  Pred. No. 0.0045;
  Matches   44;  Conservative   20;  Mismatches   44;  Indels   40;  Gaps    8;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERA--EEAKLKAKYPSL---GQKPGGSDFLMKR 57
           :::|||   |:   |||| :|:| :||     ||||  : :  ||    | || :
Db     165 eeeeeeeemekeevekqdveeeeellpvngsqeeakpqvrdfsltsssqtpgatk----- 219

Qy      58 LQKGQKYFDSGDYNMAKAKMKNKQ-----LPSAGPDKNLVT-GDHIPTPQ---------- 101
                :: ||||   :  ::::       ||| |   :|| ||   | |
Db     220 ----srhedsgdqassgvevessmgpslllpsvtp--tivtpgdqdstsqeaeatvlpaa 273

Qy     102 ------DLPQRKSSLVTSKLAG--GQVE 121
                 : ||  |   |:  ||  || |
Db     274 glgvefeapqeaseeatagaaglsgqhe 301
```

```
RESULT  15
AAW14545
ID   AAW14545 standard; Protein; 200 AA.
XX
AC   AAW14545;
XX
DT   28-OCT-1997  (first entry)
XX
DE   Streptococcus pneumoniae PspA N-terminal end.
XX
KW   PspA; pneumococcal surface protein; vaccine; otitis media;
KW   meningitis; bacteraemia; pneumonia.
XX
OS   Streptococcus pneumoniae strain Db16aapro.
XX
PN   WO9709994-A1.
XX
PD   20-MAR-1997.
XX
PF   16-SEP-1996;   96WO-US14819.
XX
PR   15-SEP-1995;   95US-0529055.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  Brooks-Walter A,  Crain MJ,  Hollingshead S;
PI   McDaniel LS, Swiatlo E,  Tart R,  Yother J;
XX
DR   WPI; 1997-202002/18.
XX
PT   Streptococcus pneumoniae surface protein PspC and truncated PspA -
PT   used in vaccines for protecting animals against S.pneumoniae
PT   infection
XX
PS   Example 6; Fig 13; 296pp; English.
XX
CC   This sequence shows the N-terminal end of pneumococcal surface
CC   protein A (PspA) of Streptococcus pneumoniae strain Db16aapro.
CC   Comparison of the N-terminal and central regions (AAW14533-57 and
CC   AAW14562-91) of PspA polypeptides from different pneumococcal strains
CC   can be used to divide the strains into several families based on
CC   sequence homologies.  PspA polypeptides, or fragments of them, can
CC   be used in vaccines to protect animals against S. pneumoniae
CC   infection and hence for the prevention of diseases such as otitis
CC   media, meningitis, bacteraemia and pneumonia.  The sequence of the
CC   3' half of the PspA alpha-helical region and the immediate 5' tip
CC   of the coding sequence are likely to be the critical sequences for
CC   predicting PspA cross-reactions and vaccine composition.
XX
SQ   Sequence   200 AA;
```

```
  Query Match             16.7%;  Score 104.5;  DB 18;  Length 200;
  Best Local Similarity   28.8%;  Pred. No. 0.0027;
  Matches   36;  Conservative   26;  Mismatches   44;  Indels   19;  Gaps    6;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           |:: || |::  || :  |||      | ||   |  : || |:    |: :|     :|
Db      50 krdaen-akkaleeakraqekyadyqrrieekaakethasleqqeankdyqlkl----kk 104

Qy      64 YFDSGDYN--------MAKAKMKNKQLPSAGPDKNLVTGD-HIPTPQDL--PQRKSSLVT 112
           | |  : :         | :|: |:|: |:   : | :  :  :| ||:|   :||| :
Db     105 yldgrnlsnssvlkkemeeaekkdkekpa---efnkirreivvpnpqelemarrksevak 161

Qy     113 SKLAG 117
           :| :|
Db     162 tkesg 166
```

```
RESULT  16
AAY31741
ID   AAY31741 standard; Protein; 605 AA.
XX
AC   AAY31741;
XX
DT   22-NOV-1999  (first entry)
XX
DE   Human podocalyxin-like sialomucin PCLP-2.
XX
KW   PCLP-2; podocalyxin-like protein; selectin; ligand; human;
KW   inflammation; autoimmune disease; tissue rejection; therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..32
FT                   /note= "signal peptide"
FT   Protein         33..605
FT                   /note= "mature protein"
XX
PN   WO9941363-A1.
XX
PD   19-AUG-1999.
XX
PF   28-JAN-1999;   99WO-US01780.
XX
```

```
PR   10-DEC-1998;   98US-0111663.
PR   11-FEB-1998;   98US-0074389.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Rosen SD,  Sassetti CM;
XX
DR   WPI; 1999-540131/45.
DR   N-PSDB; AAX87926.
XX
PT   Podocalyxin-like protein, e.g. useful for diagnosing and treating
PT   diseases associated with selectin mediated binding events
XX
PS   Claim 7; Fig 1; 63pp; English.
XX
CC   The present sequence represents a new podocalyxin-like protein
CC   (PCLP-2) having selectin ligand activity.  PCLP-2 is a sialomucin
CC   having 4 distinct regions: a mucin domain, a cysteine-rich domain,
CC   a transmembrane domain; and a cytoplasmic domain.  Its N-terminal
CC   domain differentiates it from PCLP.  Homology is shown to CD34 in
CC   the cyoplasmic tail domain.  PCLP-2 polypeptides and nucleic acids
CC   are useful for diagnosing and treating (via gene therapy) diseases
CC   associated with selectin mediated binding events, including acute
CC   and chronic inflammation, autoimmune diseases and tissue rejection,
CC   and are useful in assays for screening for therapeutic agents.
XX
SQ   Sequence   605 AA;


  Query Match             16.3%;  Score 102;  DB 20;  Length 605;
  Best Local Similarity   27.9%;  Pred. No. 0.02;
  Matches   41;  Conservative   21;  Mismatches   47;  Indels   38;  Gaps    7;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERA--EEAKLKAKYPSL---GQKPGGSDFLMKR 57
           :::|||   |:   |||: :|:| :||     |||| : :  ||    | || :
Db     165 eeeeeeeerekeevekqeeeeeeellpvngsqeeakpqvrdfsltsssqtpgatk----- 219

Qy      58 LQKGQKYFDSGDYNMAKAKMKNKQ-----LPSAGPDKNLVTGDHIPTPQ---------- 101
             :: ||||   :  ::::       ||| |   :  ||   | |
Db     220 ----srhedsgdqassgvevessmgpslllpsvtp-ttvtpgdqdstsqeaeatvlpaag 274

Qy     102 -----DLPQRKSSLVTSKLAG--GQVE 121
                 : ||  |   |:  ||  || |
Db     275 lgvefeapqeaseeatagaaglsgqhe 301


RESULT  17
AAU12251
ID   AAU12251 standard; Protein; 605 AA.
XX
AC   AAU12251;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human PRO3742 polypeptide sequence.
XX
KW   Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW   breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW   cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW   adipocyte; A-peptide; factor VIIA; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200140466-A2.
XX
PD   07-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-US32678.
XX
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28564.
PR   02-DEC-1999;   99WO-US28565.
PR   09-DEC-1999;   99US-0170262.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   20-DEC-1999;   99WO-US30999.
PR   30-DEC-1999;   99WO-US31243.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   01-MAR-2000; 2000WO-US05601.
PR   20-MAR-2000; 2000WO-US07377.
PR   21-MAR-2000; 2000WO-US07532.
PR   30-MAR-2000; 2000WO-US08439.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   10-NOV-2000; 2000WO-US30873.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Beresini M,  Deforge L,  Desnoyers L,  Filvaroff E,  Gao W;
PI   Gerritsen ME,  Goddard A,  Godowski PJ,  Gurney AL,  Sherwood S;
PI   Smith V,  Stewart TA,  Tumas D,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2001-408281/43.
DR   N-PSDB; AAS21323.
XX
PT   Isolated, secretory and transmembrane PRO polypeptide used to detect
PT   other PRO polypeptides, link bioactive molecules to cells expressing
PT   PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT   lung, breast, prostate, cervical  -
XX
PS   Claim 12; Fig 160; 813pp; English.
XX
CC   AAU12172-AAU12446 represent novel human secretory and transmembrane
CC   PRO polypeptides. The PRO polypeptides are useful to detect other
CC   PRO polypeptides, to link bioactive molecules to cells expressing
CC   PRO polypeptides, to modulate biological activities of cells expressing
CC   PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC   breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC   polypeptide expression in a cell sample to that in a control sample.
CC   Some of the 275 sequences are also useful to stimulate the release of
CC   tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC   proliferation or differentiation of chondrocytes, the proliferation or
CC   gene expression in pericyte cells, the release of proteoglycans from
CC   cartilage, the proliferation of inner ear utricular supporting cells or
CC   of T-lymphocytes, the release of a cytokine from peripheral blood
CC   monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC   the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC   skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
CC   to factor VIIA. The PRO polypeptides can be used in assays to identify
CC   molecules involved in binding interactions. The polynucleotides encoding
CC   PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC   transgenic or knock out animals and can be used in gene therapy.
XX
SQ   Sequence   605 AA;


  Query Match             16.3%;  Score 102;  DB 22;  Length 605;
  Best Local Similarity   27.9%;  Pred. No. 0.02;
  Matches   41;  Conservative   21;  Mismatches   47;  Indels   38;  Gaps    7;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERA--EEAKLKAKYPSL---GQKPGGSDFLMKR 57
           :::|||   |:   |||: :|:| :||     |||| : :  ||    | || :
Db     165 eeeeeeeerekeevekqeeeeeeellpvngsqeeakpqvrdfsltsssqtpgatk----- 219

Qy      58 LQKGQKYFDSGDYNMAKAKMKNKQ-----LPSAGPDKNLVTGDHIPTPQ---------- 101
```

```
              :: ||||   :  ::::      |||  |   :  ||   | |
Db    220 ----srhedsgdqassgvevessmgpslllpsvtp-ttvtpgdqdstsqeaeatvlpaag 274

Qy    102 -----DLPQRKSSLVTSKLAG--GQVE 121
               : ||  |   |:  ||  || |
Db    275 lgvefeapqeaseeatagaaglsgqhe 301
```

RESULT 18
AAE03767
ID AAE03767 standard; Protein; 605 AA.
XX
AC AAE03767;
XX
DT 07-AUG-2001 (first entry)
XX
DE Human gene 4 encoded secreted protein HISEL50, SEQ ID NO:37.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW endocrine disorder; infection; wound healing; vulnerary;
KW cell culture; chemotaxis; food additive;
KW binding partner identification.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..33
FT /label= signal_peptide
FT Protein 34..605
FT /note= "Mature secreted protein"
XX
PN WO200132837-A1.
XX
PD 10-MAY-2001.
XX
PF 17-OCT-2000; 2000WO-US28664.
XX
PR 02-NOV-1999; 99US-0163085.
PR 17-DEC-1999; 99US-0172411.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Fiscella M, Wei P, Lafleur DW, Olsen HS, Baker K, Ebner R;
PI Komatsoulis G, Rosen CA, Ruben SM, Duan RD, Young PE, Florence KA;
PI Moore PA, Birse CE, Ni J, Soppet DR, Shi Y;
XX
DR WPI; 2001-328782/34.
DR N-PSDB; AAD08194.
XX
PT Novel human secreted proteins and nucleic acids for diagnosing,
PT preventing and treating neurological, cardiovascular, infectious,
PT autoimmune, gastrointestinal, bone disorders, cancer, particularly
PT ovarian cancer -
XX
PS Claim 11; Page 390-392; 421pp; English.
XX
CC AAD08191-AAD08213 represent cDNAs corresponding to 19 human secreted
CC protein genes and AAE03764-AAE03786 represent the proteins they encode.
CC AAE03787-AAE03800 represent human secreted protein fragments or variants.
CC The genes and their secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 19 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein of the invention.
XX
SQ Sequence 605 AA;

Query Match 16.3%; Score 102; DB 22; Length 605;
Best Local Similarity 27.9%; Pred. No. 0.02;
Matches 41; Conservative 21; Mismatches 47; Indels 38; Gaps 7;

```
Qy      3 QKQEEENPAEETGEEKQDTQEKEGILPERA--EEAKLKAKYPSL---GQKPGGSDFLMKR 57
          :::|||   |:   |||: :|:| :||     ||||  : :  ||    | || :
Db    165 eeeeeeeerekeevekqeeeeeeellpvngsqeeakpqvrdfsltsssqtpgatk----- 219

Qy     58 LQKGQKYFDSGDYNMAKAKMKNKQ-----LPSAGPDKNLVTGDHIPTPQ---------- 101
               :: ||||   :  ::::       |||  |   :  ||   | |
Db    220 ----srhedsgdqassgvevessmgpslllpsvtp-ttvtpgdqdstsqeaeatvlpaag 274

Qy    102 -----DLPQRKSSLVTSKLAG--GQVE 121
               : ||  |   |:  ||  || |
Db    275 lgvefeapqeaseeatagaaglsgqhe 301
```

RESULT 19
AAW14547
ID AAW14547 standard; Protein; 211 AA.
XX
AC AAW14547;
XX
DT 28-OCT-1997 (first entry)
XX
DE Streptococcus pneumoniae PspA N-terminal end.
XX
KW PspA; pneumococcal surface protein; vaccine; otitis media;
KW meningitis; bacteraemia; pneumonia.
XX
OS Streptococcus pneumoniae strain Ef10197pro.
XX
FH Key Location/Qualifiers
FT Misc-difference 209
FT /note= "unidentified amino acid"
XX
PN WO9709994-A1.
XX
PD 20-MAR-1997.
XX
PF 16-SEP-1996; 96WO-US14819.
XX
PR 15-SEP-1995; 95US-0529055.
XX
PA (UABR-) UAB RES FOUND.
XX
PI Briles DE, Brooks-Walter A, Crain MJ, Hollingshead S;
PI McDaniel LS, Swiatlo E, Tart R, Yother J;
XX
DR WPI; 1997-202002/18.

```
XX
PT   Streptococcus pneumoniae surface protein PspC and truncated PspA -
PT   used in vaccines for protecting animals against S.pneumoniae
PT   infection
XX
PS   Example 6; Fig 13; 296pp; English.
XX
CC   This sequence shows the N-terminal end of pneumococcal surface
CC   protein A (PspA) of Streptococcus pneumoniae strain Ef10197pro.
CC   Comparison of the N-terminal and central regions (AAW14533-57 and
CC   AAW14562-91) of PspA polypeptides from different pneumococcal strains
CC   can be used to divide the strains into several families based on
CC   sequence homologies.  PspA polypeptides, or fragments of them, can
CC   be used in vaccines to protect animals against S. pneumoniae
CC   infection and hence for the prevention of diseases such as otitis
CC   media, meningitis, bacteraemia and pneumonia.  The sequence of the
CC   3' half of the PspA alpha-helical region and the immediate 5' tip
CC   of the coding sequence are likely to be the critical sequences for
CC   predicting PspA cross-reactions and vaccine composition.
XX
SQ   Sequence   211 AA;


  Query Match             15.6%;  Score 97.5;  DB 18;  Length 211;
  Best Local Similarity   29.4%;  Pred. No. 0.016;
  Matches   37;  Conservative   20;  Mismatches   48;  Indels   21;  Gaps    5;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           |:: || |::  || :  |||      | ||   | :  || |:    |: :|     :|
Db      51 krdaen-akkaleeakraqekyadyqrrieekaakeqqasleqqeamkdyqlkl----kk 105

Qy      64 YFD----------SGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDL--PQRKSSLV 111
           | |           :   |: | | ||       : :|    :| ||:|   :||| :|
Db     106 yldgrnlsnssvlkkemeeaekkdkekqaefnkirreiv----vpnpqelemarrksevv 161

Qy     112 TSKLAG 117
             :| :|
Db     162 kakesg 167


RESULT  20
AAW14542
ID   AAW14542 standard; Protein; 211 AA.
XX
AC   AAW14542;
XX
DT   28-OCT-1997  (first entry)
XX
DE   Streptococcus pneumoniae PspA N-terminal end.
XX
KW   PspA; pneumococcal surface protein; vaccine; otitis media;
KW   meningitis; bacteraemia; pneumonia.
XX
OS   Streptococcus pneumoniae strain Bg9739pro.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 197
FT                   /note= "unidentified amino acid"
FT   Misc-difference 198
FT                   /note= "unidentified amino acid"
XX
PN   WO9709994-A1.
XX
PD   20-MAR-1997.
XX
PF   16-SEP-1996;   96WO-US14819.
XX
PR   15-SEP-1995;   95US-0529055.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  Brooks-Walter A,  Crain MJ,  Hollingshead S;
PI   McDaniel LS, Swiatlo E,  Tart R,  Yother J;
XX
DR   WPI; 1997-202002/18.
XX
PT   Streptococcus pneumoniae surface protein PspC and truncated PspA -
PT   used in vaccines for protecting animals against S.pneumoniae
PT   infection
XX
PS   Example 6; Fig 13; 296pp; English.
XX
CC   This sequence shows the N-terminal end of pneumococcal surface
CC   protein A (PspA) of Streptococcus pneumoniae strain Bg9739pro.
CC   Comparison of the N-terminal and central regions (AAW14533-57 and
CC   AAW14562-91) of PspA polypeptides from different pneumococcal strains
CC   can be used to divide the strains into several families based on
CC   sequence homologies.  PspA polypeptides, or fragments of them, can
CC   be used in vaccines to protect animals against S. pneumoniae
CC   infection and hence for the prevention of diseases such as otitis
CC   media, meningitis, bacteraemia and pneumonia.  The sequence of the
CC   3' half of the PspA alpha-helical region and the immediate 5' tip
CC   of the coding sequence are likely to be the critical sequences for
CC   predicting PspA cross-reactions and vaccine composition.
XX
SQ   Sequence   211 AA;


  Query Match             15.1%;  Score 94.5;  DB 18;  Length 211;
  Best Local Similarity   28.0%;  Pred. No. 0.035;
  Matches   35;  Conservative   26;  Mismatches   45;  Indels   19;  Gaps    6;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQK 63
           |:: || |::  || :  |||      | ||   |    || |:    |: :|     :|
Db      50 krdaen-akkaleeakraqekyadyqrrieekaaketqasleqqeankdyqlkl----kk 104

Qy      64 YFDSGDYN--------MAKAKMKNKQLPSAGPDKNLVTGD-HIPTPQDL--PQRKSSLVT 112
           | |  : :         | :|: |:|:  :   ::| :  :  :| ||:|   :||| :|
Db     105 yldgrnlsnssvlkkemeeaekkdkenqa---eenkirreivvpnpqelemarrksevvk 161

Qy     113 SKLAG 117
           :  :|
Db     162 atesg 166


RESULT  21
AAM93868
ID   AAM93868 standard; Protein; 360 AA.
XX
AC   AAM93868;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide, SEQ ID NO: 3973.
XX
KW   Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS   Homo sapiens.
XX
PN   EP1130094-A2.
XX
PD   05-SEP-2001.
XX
PF   07-JUL-2000; 2000EP-0114089.
XX
PR   08-JUL-1999;   99JP-0194486.
PR   11-JAN-2000; 2000JP-0118774.
PR   02-MAY-2000; 2000JP-0183765.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
PI   Wakamatsu A,  Sugiyama T,  Nagai K,  Kojima S,  Otsuki T,  Koga H;
XX
DR   WPI; 2001-524255/58.
DR   N-PSDB; AAK94826.
```

```
XX
PT   830 Primers useful for synthesizing full length cDNA clones and their
PT   use in genetic manipulation -
XX
PS   Claim 8; SEQ ID NO 3973; 1380pp + sequence listing; English.
XX
CC   The invention relates to primers for synthesising full length cDNA
CC   clones. 830 cDNA molecules encoding a human protein have been
CC   isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC   molecules have been determined. Primers for synthesising the full length
CC   cDNA are useful for clarifying the function of the protein encoded by
CC   the cDNA. The full length clones were obtained by construction of full
CC   length enriched cDNA libraries that were synthesised by the oligo-capping
CC   method. The primers enable the production of the full length cDNA easily
CC   without any special methods. The present sequence is a polypeptide
CC   encoded by a full length human cDNA of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ   Sequence   360 AA;


  Query Match             14.5%;  Score 90.5;  DB 22;  Length 360;
  Best Local Similarity   26.8%;  Pred. No. 0.18;
  Matches   40;  Conservative   23;  Mismatches   51;  Indels   35;  Gaps    6;

Qy       4 KQEEENPAEETGEEKQDTQEKEGILPER------------AEEAKLKAKYP----SLGQK 47
           ||: :   ::  |||::  :|:| | |:            :|:|:|:        :  ||
Db      89 kqqnqeitdqleeekkechqkqgalreqlqvhiqtigilvsekaelqtalahtqhaarqk 148

Qy      48 PGGSDFLMKRLQKGQKYFDSGDYNMA---------KAKMKNKQLPSAGPDKNLVTGDHIP 98
             | |: |  |||  ::    |:   |          ||   ||:|        |    :
Db     149 egesedlasrlqysrrrv--geleralsavstqqkkadrynkeltkerdalrlelykntq 206

Qy      99 TPQDLPQRKSS-------LVTSKLAGGQV 120
           : :|| | ||        ||| | || |:
Db     207 snedlkqekseleeklrvlvtek-agmql 234


RESULT  22
ABG11252
ID   ABG11252 standard; Protein; 193 AA.
XX
AC   ABG11252;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #11243.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS75439.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 41611; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   193 AA;


  Query Match             14.2%;  Score 89;  DB 22;  Length 193;
  Best Local Similarity   27.3%;  Pred. No. 0.12;
  Matches   30;  Conservative   17;  Mismatches   41;  Indels   22;  Gaps    4;

Qy       1 MSQKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQK 60
           |:||:|||   :|  ||::: :|:|    |  || | | |                :: ::
Db       1 maqkkeeeeekkeeeeeeeeeeeeee--eeeeeeekkkkk--------------eekrer 44

Qy      61 GQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSL 110
             :||      |      :|| | |  |     | :| |        |:|:
Db      45 skkykqqqqqn----tkrnknlts--pqdetqtanhkpeehpeentknsv 88


RESULT  23
ABB69600
ID   ABB69600 standard; Protein; 1751 AA.
XX
AC   ABB69600;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 35592.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL13703.
XX
```

```
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 35592; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1751 AA;


  Query Match             14.0%;  Score 87.5;  DB 22;  Length 1751;
  Best Local Similarity   26.3%;  Pred. No. 2.9;
  Matches   35;  Conservative   18;  Mismatches   57;  Indels   23;  Gaps    4;

Qy       3 QKQEEENPA-------EETGEEKQDTQEKEGILPERAEEAKLKAKYPSLG---QKPGGSD 52
           |  :| |||         :|||::   :| |   :|: :   :    |: |    :
Db     976 qssdesnparieeendsnsgeereg--kkrgtsgqrstdrsaerispafgfnrlelrskq 1033

Qy      53 FLMKRLQKGQKYFDSGDYNMAKAK----------MKNKQLPSAGPDKNLVTGDHIPTPQ 101
           |  | :  |  :  ||   |  ||           :|:||    ||   || |  |  :
Db   1034 fkskpipigahhtvsgheamdlakrlneelashqqqqnqqlrpispdiravtpdceprsr 1093

Qy     102 DLPQRKSSLVTSK 114
             || :| | :|
Db   1094 sfeqrptsgvcak 1106


RESULT  24
ABG25869
ID   ABG25869 standard; Protein; 615 AA.
XX
AC   ABG25869;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #25860.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS90056.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 56228; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   615 AA;


  Query Match             13.6%;  Score 85;  DB 22;  Length 615;
  Best Local Similarity   25.0%;  Pred. No. 1.4;
  Matches   29;  Conservative   23;  Mismatches   36;  Indels   28;  Gaps    4;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQ 62
           ||||||   ||  ||::: :|:|    |  || :        |:: ||         :|:
Db     361 qkqeeeeeeeeeeeeeeeeeeeeeeeeeeeege--------geregg---------ege 403

Qy      63 KYFDSGDY-----NMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTS 113
           :  : |::     :  :|| :  :   : ||      : ||    :  |   ||::
Db     404 regeegehdrfsthkpqakdetmkyirskpd------nriperhyiagrprelvst 453


RESULT  25
ABG09291
ID   ABG09291 standard; Protein; 659 AA.
XX
AC   ABG09291;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #9282.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS73478.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
```

```
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 39650; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   659 AA;


  Query Match             13.6%;  Score 85;  DB 22;  Length 659;
  Best Local Similarity   25.0%;  Pred. No. 1.6;
  Matches   29;  Conservative   23;  Mismatches   36;  Indels   28;  Gaps    4;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKRLQKGQ 62
           ||||||   ||  ||::: :|:|    |  || :        |:: ||         :|:
Db     405 qkqeeeeeeeeeeeeeeeeeeeeeeeeeeege--------geregg---------ege 447

Qy      63 KYFDSGDY-----NMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTS 113
           :  : |::     :  :|| :  :   : ||      : ||    :  |   ||::
Db     448 regeegehdrfsthkpqakdetmkyirskpd------nriperhyiagrprelvst 497


RESULT  26
ABG04385
ID   ABG04385 standard; Protein; 131 AA.
XX
AC   ABG04385;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #4376.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS68572.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 34744; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   131 AA;


  Query Match             13.5%;  Score 84.5;  DB 22;  Length 131;
  Best Local Similarity   26.9%;  Pred. No. 0.23;
  Matches   28;  Conservative   17;  Mismatches   52;  Indels    7;  Gaps    2;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGS------DFLMK 56
           :::|||   ||  ||::: :|:|    |  || : : :   ||:  |||       |
Db      28 eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeegeeeelgrlhggsgkvrglgften 87

Qy      57 RLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTP 100
           :     :     |   :| |  |   | |  :  :| |  |  |
Db      88 qqgstnrqhqrednrkskqkkinntkpea-teslivngititap 130


RESULT  27
ABG11269
ID   ABG11269 standard; Protein; 147 AA.
XX
AC   ABG11269;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #11260.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
```

```
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS75456.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 41628; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   147 AA;


  Query Match             13.4%;  Score 84;  DB 22;  Length 147;
  Best Local Similarity   32.9%;  Pred. No. 0.3;
  Matches   23;  Conservative    8;  Mismatches   17;  Indels   22;  Gaps    2;

Qy       3 QKQEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQK---------------- 47
           :|:|||   ||  ||::: :|:|    |  || |     | ||:|
Db      76 kkeeeeeeeeeeeeeeeeeeeee--eeeeeekkhslxdpdlgekenkgqrkrklargts 133

Qy      48 -----PGGSD 52
                |||||
Db     134 qpslspggsd 143


RESULT  28
ABB68123
ID   ABB68123 standard; Protein; 969 AA.
XX
AC   ABB68123;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 31161.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL12226.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 31161; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   969 AA;


  Query Match             13.4%;  Score 84;  DB 22;  Length 969;
  Best Local Similarity   27.4%;  Pred. No. 3.2;
  Matches   32;  Conservative   20;  Mismatches   51;  Indels   14;  Gaps    5;

Qy       1 MSQKQEEEN--PAEETGEEK-QDTQEKEGILP-----ERAEEAKLKAKYPSLGQKPGGSD 52
           :| : |||    ||  |: |   |:: :|:||     :|  |         : ||
Db      27 vssqdeeetcaeaespgqrkwtkteqsdgmlplsdkfqrirellrlefqreisqkveqla 86

Qy      53 FLMKRLQKGQKYFDSGDYNMAKAKMKNKQLPSAGPD-----KNLVTGDHIPTPQDLP 104
             : :|| :|::  |   | :     : :|:|  | |      : : || |  || ||
Db      87 eidrrllqgrqlldrlryqvvseyyrkqqvpltgadiakvrgdslfgddiaapq-lp 142


RESULT  29
AAB65635
ID   AAB65635 standard; Protein; 2286 AA.
XX
AC   AAB65635;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Novel protein kinase, SEQ ID NO: 162.
XX
KW   Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
KW   immunosuppressive; cardiant; renal; antiinflammatory; antiasthmatic;
KW   dermatological; antidiabetic; antiinfertility; gene therapy; vaccine;
KW   immune disorder; cardiovascular disease; neurodegenerative disease;
KW   cancer; autoimmune disorder; stroke; inflammatory bowel disease;
KW   inflammatory pelvic disease; multiple sclerosis; psoriasis.
XX
OS   Homo sapiens.
XX
PN   WO200073469-A2.
XX
PD   07-DEC-2000.
XX
PF   26-MAY-2000; 2000WO-US14842.
XX
PR   28-MAY-1999;   99US-0136503.
XX
PA   (SUGE-) SUGEN INC.
XX
PI   Plowman GD,  Martinez R,  Whyte D,  Sudersanam S;
XX
```

DR WPI; 2001-032161/04.
DR N-PSDB; AAF44662.
XX
PT Nucleic acids encoding kinase polypeptides, useful for diagnosing and
PT treating immune-related diseases and disorders, cardiovascular disease,
PT neurodegenerative diseases and/or cancers -
XX
PS Claim 10; Fig 1; 310pp; English.
XX
CC The present sequence is a novel protein kinase. The novel protein kinases
CC and the nucleic acids that encode them may be used in the treatment and
CC diagnosis of diseases associated with inappropriate kinase expression
CC such as immune-related diseases and disorders, cardiovascular disease,
CC neurodegenerative diseases and/or cancers. The nucleic acids and
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays. The kinase polypeptides may be used as antigens in the production
CC of antibodies of kinase expression and activity. Anti-kinase antibodies
CC and kinase antagonists may also be used to down regulate kinase
CC expression and activity. Diseases related to kinase expression and
CC activity include rheumatoid arthritis, atherosclerosis, autoimmune
CC disorders, complications of organ transplantation, myocardial infarction,
CC immune disorders, cardiomyopathies, strokes, renal failure,
CC oxidative-stress related disorders, chronic inflammatory bowel disease,
CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
CC reproductive disorders.
XX
SQ Sequence 2286 AA;

Query Match 13.4%; Score 84; DB 22; Length 2286;
Best Local Similarity 27.4%; Pred. No. 9.5;
Matches 29; Conservative 16; Mismatches 43; Indels 18; Gaps 4;

```
Qy       5 QEEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKP--GGSDFLMKRLQKGQ 62
           ::|     |||       :::|  | : :||      || ||:|  | |       ::|:
Db     975 edealgtpetgaatpmdwqeqgrapsqdqeapspealpspgqepaagasp------rrge 1028

Qy      63 KYFDSGDYNMAKAKMKNKQLPSAGPDKNLVTGDHIPTPQDLPQRKS 108
                 : :       || ||| : |  | |     :||||:|
Db    1029 ---------lrrgssaesalpragp-relgrglhkaasvelpqrrs 1064
```

RESULT 30
ABB71057
ID ABB71057 standard; Protein; 281 AA.
XX
AC ABB71057;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 39963.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL15160.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 39963; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 281 AA;

Query Match 13.4%; Score 83.5; DB 22; Length 281;
Best Local Similarity 36.5%; Pred. No. 0.77;
Matches 27; Conservative 12; Mismatches 20; Indels 15; Gaps 4;

```
Qy       3 QKQEEENP-----AEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGGSDFLMKR 57
           | ||||:|     |||  ||::| :|:|   |:       :   |:| | |||   :::
Db     120 qgqeeespdqdgeaeeaeeeeedeeeqeeatpqp------qplppalpq-pgg---vlee 169

Qy      58 LQKGQKYFDSGDYN 71
           |  ||   :  |||
Db     170 laggqvdeeledyn 183
```

RESULT 31
AAG16884
ID AAG16884 standard; Protein; 374 AA.
XX
AC AAG16884;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 17697.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

```
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             13.4%;  Score 83.5;  DB 21;  Length 374;
  Best Local Similarity   42.2%;  Pred. No. 1.1;
  Matches   19;  Conservative   10;  Mismatches   15;  Indels    1;  Gaps    1;

Qy       6 EEENPAEETGEEKQDTQEKEGILPERAEEAKLKAKYPSLGQKPGG 50
           ||::  |: ||| :: :::|    |  ||:| | | ||:| | ||
Db     311 eeddidedegeedeededeeddddedeeesktkkk-psignkkgg 354


RESULT  32
AAR63437
ID   AAR63437 standard; Protein; 619 AA.
XX
AC   AAR63437;
XX
DT   19-JUL-1995  (first entry)
XX
DE   Pneumococcal surface protein A from S.pneumoniae Rx1.
XX
KW   Pneumococcal surface protein A; PspA; Streptococcus; PCR; pneumococcal;
KW   primer; protection-eliciting epitope; epitope; vaccine; amplify.
XX
OS   Streptococcus pneumoniae strain Rx1.
XX
FH   Key             Location/Qualifiers
FT   Misc_feature    192..260
FT                   /note= "protein fragment of Claim 1"
XX
PN   EP622081-A.
XX
PD   02-NOV-1994.
XX
PF   19-APR-1994;   94EP-0302767.
XX
PR   20-APR-1993;   93US-0048896.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  McDaniel LS,  Yother JL;
XX
DR   WPI; 1994-359522/45.
DR   N-PSDB; AAQ78131.
XX
PT   Epitopic regions of Pneumococcal surface protein A - derived from the
PT   Rx1 PspA strain, for the preparation of cross-reactive vaccines for
PT   the prevention of pneumococcal infections
XX
PS   Disclosure; Page 13-16; 26pp; English.
XX
CC   The aminbo acid sequence of the novel Pneumococcal surface protein A
CC   (PspA) from Streptococcus pneumoniae strain Rx1.  The gene was PCR
CC   amplified from S.pneumoniae genomic DNA using the primers AAQ78132-5.
CC   The gene was used to derive truncated peptide fragments containing
CC   protection-eliciting epitopes for use in vaccines against pneumococcal
CC   diseases.  The epitopic fragments are derived from amino acids 192-260
CC   and optionally contain a further 25 a.a. residues at both the N- and
CC   C-terminal regions of the peptide.  The epitopic peptide fragments may
CC   be derived from different strains of S.pneumoniae and are homologous to
CC   the Rx1 strain epitope.
XX
SQ   Sequence   619 AA;


  Query Match             13.3%;  Score 83;  DB 15;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 2.4;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 yntdgsmatgwlqn 427


RESULT  33
AAR87598
ID   AAR87598 standard; Protein; 619 AA.
XX
AC   AAR87598;
XX
DT   04-JUL-1996  (first entry)
XX
DE   Pneumococcal surface protein (PspA).
XX
KW   PspA; pneumococcal surface protein; truncated; immunoprotective;
KW   soluble fragment; insertion-duplication mutagenesis.
XX
OS   Streptococcus pneumoniae strain Rx1.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..31
FT                   /label= signal_peptide
FT   Protein         32..619
FT                   /label= mature_protein
FT   Region          32..319
FT                   /label= alpha-helical_coiled-coil region
FT                   /note= "contains a seven-residue periodicity"
FT   Region          320..401
FT                   /note= "proline-rich region"
FT   Region          402..421
FT                   /note= "repeat region"
FT   Region          422..441
FT                   /note= "repeat region"
FT   Region          442..461
FT                   /note= "repeat region"
FT   Region          462..481
FT                   /note= "repeat region"
FT   Region          482..501
FT                   /note= "repeat region".
FT   Region          502..521
FT                   /note= "repeat region"
FT   Region          522..541
FT                   /note= "repeat region"
FT   Region          542..561
FT                   /note= "repeat region"
FT   Region          562..581
FT                   /note= "repeat region"
FT   Region          582..619
FT                   /note= "hydrophobic region starts in last repeat
FT                          region is potential membrane-spanning region"
XX
PN   US5476929-A.
XX
PD   19-DEC-1995.
XX
```

```
PF   15-FEB-1991;   91US-0656773.
XX
PR   12-FEB-1992;   92US-0835698.
PR   15-FEB-1991;   91US-0656773.
PR   03-JUN-1993;   93US-0072070.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  Mcdaniel LS,  Yother JL;
XX
DR   WPI; 1996-049021/05.
DR   N-PSDB; AAT08979.
XX
PT   New pneumococcal surface protein A fragments - comprise proline-rich
PT   region and/or repeat region, used for detection of Streptococcus
PT   pneumoniae.
XX
PS   Claim 1; Column 15-20; 23pp; English.
XX
CC   The present sequence is that of PspA (pneumococcal surface protein A)
CC   encoded by AAT08979.  Through the technique of insertion-duplication
CC   mutagenesis of the pspA gene of the strain Rx1 of Streptococcus
CC   pneumoniae with plasmids contg. cloned fragments of the pspA structural
CC   gene, it has been possible to produce soluble fragments of PspA that are
CC   secreted by pneumococci.  The method can be used to provide an
CC   immunoprotective truncated PspA protein. Primers and probes based on the
CC   present sequence are claimed, and are useful for the detection of (at
CC   least 32) S. pneumoniae strains.
XX
SQ   Sequence   619 AA;


  Query Match             13.3%;  Score 83;  DB 17;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 2.4;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy      66 DSGDYNMAKAKMKN 79
           : | :||   ::|
Db     414 yntdgsmatgwlqn 427


RESULT  34
AAR86911
ID   AAR86911 standard; Protein; 619 AA.
XX
AC   AAR86911;
XX
DT   11-MAY-1996  (first entry)
XX
DE   Pneumococcal surface protein A.
XX
KW   Pneumococcal surface protein A; PspA; cross-protection; vaccine;
KW   Streptococcus pneumoniae; probe; primer; polymerase chain reaction;
KW   otitis media; meningitis, bacteraemia; pneumonia; epitope.
XX
OS   Streptococcus pneumoniae strain Rx1.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..31
FT                   /label= Sig_peptide
FT   Region          32..288
FT                   /note= "N-terminal region is highly charged and
FT                          includes an alpha-helix structure"
FT   Region          289..619
FT                   /note= "C-terminal region includes a proline-rich
FT                          region and a repeat region"
XX
PN   AU9520112-A.
XX
PD   30-NOV-1995.
XX
PF   18-MAY-1995;   95AU-0020112.
XX
PR   07-OCT-1994;   94US-0319795.
PR   20-MAY-1994;   94US-0246636.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  McDaniel LS,  Yother JL;
XX
DR   WPI; 1996-030801/04.
DR   N-PSDB; AAT07178.
XX
PT   Pneumococcal DNA primers and probes - amplify and detect
PT   cross-protective epitope(s) from Streptococcus pneumoniae surface
PT   protein A
XX
PS   Disclosure; Page 41-43; 61pp; English.
XX
CC   Surface protein A, PspA (AAR86911), of Streptococcus pneumoniae Rx1
CC   is the product of the pspA gene (AAT07178).  PspA includes regions,
CC   comprising e.g. amino acids 182-588, 293-588 and 192-299, that elicit
CC   cross-protection against challenge by multiple wild-type strains of
CC   S. pneumoniae.  These cross-reactive epitopes can be prepd. by
CC   expression of DNA obtd. by PCR amplification (see AAT07179-96), for
CC   use in vaccine compsns.
XX
SQ   Sequence   619 AA;


  Query Match             13.3%;  Score 83;  DB 17;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 2.4;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy      66 DSGDYNMAKAKMKN 79
           : | :||   ::|
Db     414 yntdgsmatgwlqn 427


RESULT  35
AAY41838
ID   AAY41838 standard; Protein; 619 AA.
XX
AC   AAY41838;
XX
DT   08-DEC-1999  (first entry)
XX
DE   Streptococcus pneumoniae Rx1 PspA protein sequence.
XX
KW   Streptococcus pneumoniae Rx1; PspA; immunoprotective; vaccine;
KW   diagnosis; infection; pneumococcal surface protein A.
XX
OS   Streptococcus pneumoniae.
XX
PN   US5965400-A.
XX
PD   12-OCT-1999.
XX
PF   23-MAY-1994;   94US-0247491.
XX
PR   12-FEB-1992;   92US-0835698.
PR   15-FEB-1991;   91US-0656773.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Yother JL,  Briles DE;
XX
DR   WPI; 1999-579913/49.
```

```
DR   N-PSDB; AAZ25063.
XX
PT   DNA encoding a truncated pneumococcal surface protein A used in the
PT   development of pneumococcal infections -
XX
PS   Claim 1; Fig 3; 27pp; English.
XX
CC   The present sequence represents Streptococcus pneumoniae Rx1
CC   immunoprotective Pneumococcal surface protein A (PspA). The present
CC   invention also describes a method of forming the immunoprotective
CC   truncated PspA, comprising incorporating a vector comprising the
CC   isolated DNA molecule encoding PspA (I), into a bacterium via
CC   transformation. (I) is used to design primers which are capable of
CC   detecting a large number of S. pneumoniae strains, which in turn can
CC   be used to diagnose pneumococcal infection in mammals (e.g. humans),
CC   independent of the strain which has caused it. The PspA protein is
CC   used to develop a vaccine against pneumococcal infection comprising,
CC   as an immunologically-active component, a live attenuated or killed
CC   bacteria containing a gene coding for the truncated form of PspA.
XX
SQ   Sequence   619 AA;


  Query Match             13.3%;  Score 83;  DB 20;  Length 619;
  Best Local Similarity   31.1%;  Pred. No. 2.4;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 yntdgsmatgwlqn 427


RESULT  36
AAW70336
ID   AAW70336 standard; Protein; 648 AA.
XX
AC   AAW70336;
XX
DT   18-NOV-1998  (first entry)
XX
DE   Pneumococcal surface protein A (PspA).
XX
KW   Pneumococcal surface protein A gene; PspA; PspA epitope; vaccine;
KW   insertion-duplication mutagenesis; otitis media; meningitis;
KW   bacteraemia; pneumonia.
XX
OS   Streptococcus pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..31
FT                   /note= "Signal peptide"
FT   Protein         32..648
FT                   /note= "PspA"
FT   Region          32..319
FT                   /note= "alpha-helical coil region representing the
FT                         truncated PspA of the invention "
FT   misc_difference 647
FT                   /note= "Encoded by AGG"
XX
PN   US5804193-A.
XX
PD   08-SEP-1998.
XX
PF   17-MAR-1994;   94US-0214222.
XX
PR   12-FEB-1992;   92US-0835698.
PR   15-FEB-1991;   91US-0656773.
PR   17-MAR-1994;   94US-0214222.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  Yother JL;
XX
DR   WPI; 1998-505588/43.
DR   N-PSDB; AAV33264.
XX
PT   Truncated pneumococcal surface protein - useful in vaccines against
PT   pneumococcal infection
XX
PS   Example 3; Fig 3A-3C; 22pp; English.
XX
CC   The present sequence represents the Streptococcus pneumoniae Rx1
CC   pneumococcal surface protein A (PspA).  The invention provides a
CC   purified truncated form of PspA, formed by an insertion-duplication
CC   mutagenesis technique, comprising of the first 288 N-terminal residues of
CC   the mature form of wild-type PspA (AAW70336).  The truncated PspA
CC   contains immunoprotective epitopes of PspA.  The invention claims for a
CC   vaccine against pneumococcal infection, comprising live-attenuated or
CC   killed S. pneumoniae, containing the gene coding for the truncated PspA
CC   protein. The truncated protein, optionally conjugated to a poorly
CC   immunogenic or nonimmunogenic molecule, is claimed to be useful in
CC   vaccines against pneumococcal infection, especially otitis media,
CC   meningitis, bacteraemia and pneumonia.
XX
SQ   Sequence   648 AA;


  Query Match             13.3%;  Score 83;  DB 19;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 2.5;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 yntdgsmatgwlqn 427


RESULT  37
AAW62274
ID   AAW62274 standard; Protein; 648 AA.
XX
AC   AAW62274;
XX
DT   22-SEP-1998  (first entry)
XX
DE   Streptococcus pneumoniae pspA protein.
XX
KW   Streptococcus pneumoniae strain Rx1; pspA; immunoprotective; immunogen;
KW   pneumococcal surface protein A; cholera toxin B subunit; fusion protein;
KW   antigenic.
XX
OS   Streptococcus pneumoniae.
XX
PN   US5753463-A.
XX
PD   19-MAY-1998.
XX
PF   06-JUN-1995;   95US-0469434.
XX
PR   12-FEB-1992;   92US-0835698.
PR   15-FEB-1991;   91US-0656773.
PR   03-JUN-1993;   93US-0072065.
PR   06-JUN-1995;   95US-0469434.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  Yother JL;
XX
DR   WPI; 1998-311399/27.
```

```
DR   N-PSDB; AAV39470.
XX
PT   Truncated pneumococcal surface protein and cholera toxin B sub-unit
PT   fusion protein - useful as an immunogen against Streptococcus
PT   pneumoniae
XX
PS   Claim 1; Fig 3; 22pp; English.
XX
CC   The present sequence represents the pneumococcal surface protein A (PspA)
CC   protein from Streptococcus pneumoniae. A recombinant DNA molecule has
CC   been developed which encodes a fusion protein comprising a truncated
CC   form of PspA and cholera toxin B subunit (CTB), where the DNA molecule
CC   comprises a nucleotide sequence encoding the truncated PspA linked by
CC   an in-frame genetic fusion to a ctxB gene, and where the truncated PspA
CC   contains immunoprotective epitopes and up to 90% of the whole PspA
CC   protein, except for the cell membrane anchor region. The fusion protein
CC   is useful for providing an immunogen to protect neonates and children
CC   against S.pneumoniae. Most antigenic proteins of this strain are not
CC   immunogenic enough to provide protection. The antigenic epitopes of the
CC   fusion protein are directed against capsular polysaccharide antigens of
CC   S.pneumoniae, specifically it contains the protective epitopes of PspA.
CC   The protein can also be used in solid-phase immunoadsorbent assays,
CC   since it is readily bound to supports coated with monosialoganglioside
CC   GM1. The fusion protein is more immunogenic against S.pneumoniae than
CC   using PspA alone as the immunogen.
XX
SQ   Sequence   648 AA;


  Query Match             13.3%;  Score 83;  DB 19;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 2.5;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 yntdgsmatgwlqn 427


RESULT  38
AAY41837
ID   AAY41837 standard; Protein; 648 AA.
XX
AC   AAY41837;
XX
DT   08-DEC-1999  (first entry)
XX
DE   Streptococcus pneumoniae Rx1 PspA protein sequence.
XX
KW   Streptococcus pneumoniae Rx1; PspA; immunoprotective; vaccine;
KW   diagnosis; infection; pneumococcal surface protein A.
XX
OS   Streptococcus pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 619..620
FT                   /note= "a stop codon is present in the nucleotide
FT                         sequence at this position"
FT   Misc-difference 621..622
FT                   /note= "a stop codon is present in the nucleotide
FT                         sequence at this position"
FT   Misc-difference 625..626
FT                   /note= "a stop codon is present in the nucleotide
FT                         sequence at this position"
FT   Misc-difference 630..631
FT                   /note= "a stop codon is present in the nucleotide
FT                         sequence at this position"
FT   Misc-difference 632..633
FT                   /note= "a stop codon is present in the nucleotide
FT                         sequence at this position"
XX
PN   US5965400-A.
XX
PD   12-OCT-1999.
XX
PF   23-MAY-1994;   94US-0247491.
XX
PR   12-FEB-1992;   92US-0835698.
PR   15-FEB-1991;   91US-0656773.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Yother JL,  Briles DE;
XX
DR   WPI; 1999-579913/49.
DR   N-PSDB; AAZ25063.
XX
PT   DNA encoding a truncated pneumococcal surface protein A used in the
PT   development of pneumococcal infections -
XX
PS   Claim 1; Fig 3; 27pp; English.
XX
CC   The present sequence represents Streptococcus pneumoniae Rx1
CC   immunoprotective Pneumococcal surface protein A (PspA). The present
CC   invention also describes a method of forming the immunoprotective
CC   truncated PspA, comprising incorporating a vector comprising the
CC   isolated DNA molecule encoding PspA (I), into a bacterium via
CC   transformation. (I) is used to design primers which are capable of
CC   detecting a large number of S. pneumoniae strains, which in turn can
CC   be used to diagnose pneumococcal infection in mammals (e.g. humans),
CC   independent of the strain which has caused it. The PspA protein is
CC   used to develop a vaccine against pneumococcal infection comprising,
CC   as an immunologically-active component, a live attenuated or killed
CC   bacteria containing a gene coding for the truncated form of PspA.
XX
SQ   Sequence   648 AA;


  Query Match             13.3%;  Score 83;  DB 20;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 2.5;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :||  :     : : |  ||
Db     354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 yntdgsmatgwlqn 427


RESULT  39
AAW87879
ID   AAW87879 standard; Protein; 648 AA.
XX
AC   AAW87879;
XX
DT   19-MAR-1999  (first entry)
XX
DE   A pneumococcal surface protein (PspA).
XX
KW   Pneumococcal surface protein; PspA; Streptococcus pneumoniae Rx1;
KW   truncated protein; epitope-containing region; vaccine;
KW   pneumococcal infection.
XX
OS   Streptococcus pneumoniae.
XX
PN   US5856170-A.
XX
PD   05-JAN-1999.
XX
PF   06-JUN-1995;   95US-0467852.
XX
```

```
PR   12-FEB-1992;   92US-0835698.
PR   15-FEB-1991;   91US-0656773.
PR   23-MAY-1994;   94US-0247491.
PR   06-JUN-1995;   95US-0467852.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  Yother JL;
XX
DR   WPI; 1999-105118/09.
DR   N-PSDB; AAV84069.
XX
PT   DNA encoding truncated pneumococcal PspA protein - useful for
PT   producing recombinant truncated protein
XX
PS   Disclosure; Fig 3A-D; 27pp; English.
XX
CC   The present sequence represents a pneumococcal surface protein (PspA)
CC   of Streptococcus pneumoniae Rx1. The specification describes
CC   truncated forms of PspA which contain at least the N-terminal
CC   alpha-helical protective epitope-containing region. The truncated
CC   PspA proteins can be used in vaccines against pneumococcal
CC   infections.
XX
SQ   Sequence   648 AA;
```

```
  Query Match             13.3%;  Score 83;  DB 20;  Length 648;
  Best Local Similarity   31.1%;  Pred. No. 2.5;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 yntdgsmatgwlqn 427
```

```
RESULT  40
AAW92456
ID   AAW92456 standard; Protein; 653 AA.
XX
AC   AAW92456;
XX
DT   21-APR-1999  (first entry)
XX
DE   S. pneumoniae truncated pspA protein.
XX
KW   PspA; pneumococcal surface protein A; immunoprotection; detection; GM1;
KW   solid phase immunosorbent assay; epitope; cell membrane anchor region;
KW   cholera toxin B subunit; CTB; monosialoganglioside; fusion protein.
XX
OS   Streptococcus pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 620
FT                   /note= "in frame stop codon encoded by TAA"
FT   Misc-difference 623
FT                   /note= "In frame stop codon encoded by TAA"
FT   Misc-difference 628
FT                   /note= "In frame stop codon encoded by TAA"
FT   Misc-difference 634
FT                   /note= "in frame stop codon encoded by TAA"
FT   Misc-difference 636
FT                   /note= "In frame stop codon encoded by TGA"
XX
PN   US5871943-A.
XX
PD   16-FEB-1999.
XX
PF   06-JUN-1995;   95US-0468718.
```

```
XX
PR   12-FEB-1992;   92US-0835698.
PR   15-FEB-1991;   91US-0656773.
PR   03-JUN-1993;   93US-0072068.
PR   06-JUN-1995;   95US-0468718.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Briles DE,  Yother JL;
XX
DR   WPI; 1999-166635/14.
DR   N-PSDB; AAX02012.
XX
PT   Immunosorbent assay for pneumococcal surface protein A antigen or
PT   antibody  - for diagnosis of infection by Streptococcus pneumoniae
XX
PS   Claim 1; Fig 3A-C; 24pp; English.
XX
CC   This sequence represents a truncated form of the Streptococcus pneumoniae
CC   PspA protein which is used in a solid phase immunosorbent assay for
CC   detecting a PspA (pneumococcal surface protein A) antibody and antigen.
CC   This truncated protein contains the immunoprotective epitopes of the
CC   complete protein (up to 90% of PspA but excludes the cell-membrane
CC   anchor region) fused to the B subunit of cholera toxin (CTB) which is
CC   bound to monosialoganglioside (GM1) coated on the substrate. The use of
CC   a fusion between truncated PspA and cholera toxin B subunit (CTB)
CC   allows the support to be coated without having to isolate PspA fragments
CC   since CTB binds specifically to the GM1 coating the solid support.
XX
SQ   Sequence   653 AA;
```

```
  Query Match             13.3%;  Score 83;  DB 20;  Length 653;
  Best Local Similarity   31.1%;  Pred. No. 2.5;
  Matches   23;  Conservative   13;  Mismatches   36;  Indels    2;  Gaps    1;

Qy       8 ENPAEETGEEKQDTQEKEGILPERAEEA--KLKAKYPSLGQKPGGSDFLMKRLQKGQKYF 65
           | |||:   || | |: |     |:||   :|  : |   :|| :     : : |  ||
Db     354 ekpaeqpkpektddqqaeedyarrseeeynrltqqqppkaekpapapktgwkqengmwyf 413

Qy      66 DSGDYNMAKAKMKN 79
             : | :||   ::|
Db     414 yntdgsmatgwlqn 427


Search completed: September  7, 2002, 10:45:23
Job time: 154 sec
```